Crystallographic Studies of NAD^+-Dependent L- and D- 2-Hydroxyacid Dehydrogenases

A thesis submitted for the degree of Doctor of Philosophy at the University of Bristol in the Faculty of Science

by

Balvinder Dhaliwal

March 2001

Department of Biochemistry

University of Bristol

Abstract

D-Hydroxyisocaproate Dehydrogenase: A Broad Specificity D-2-Hydroxyacid Dehydrogenase

L.bulgaricus D-2-hydroxyisocaproate dehydrogenase (D-HicDH) catalyses the NAD^+-dependent stereospecific and reversible reduction of a broad spectrum of aliphatic and aromatic 2-ketoacids.

Numerous crystal forms of the wild type and H205Q mutant D-HicDH were found. However, attempts at phasing the subsequent diffraction data using several molecular replacement programs were unsuccessful.

The multiple isomorphous replacement method was then used to try and phase the D-HicDH data sets. Unfortunately, suitable heavy atom derivatives were not found.

The multiwavelength anomalous dispersion phasing method was then employed to try and solve the D-HicDH structure. In order to obtain a usable anomalous signal, methionine residues in the protein had to be substituted for selenomethionine. 100% Selenomethionyl incorporation into the protein was eventually achieved by transforming the plasmid harbouring the D-HicDH gene into a methionine auxotrophic cell strain. However, the subsequent Se-Met D-HicDH were not of sufficient quality for MAD phasing experiments.

*Pf*LDH Stability Studies

The L-lactate dehydrogenase from *Plasmodium falciparum* is a target for antimalarial drug design. Previous kinetic experiments had shown the enzyme to be catalytically unstable against oxidation. This was thought to be the result of inappropriate inter-subunit disulphide bond formation. In order to stabilise the enzyme, one of the cysteine residues at the subunit interface was mutated to valine, alanine and serine.

The mutant crystal structures reveal that the initial enzymatic rate is dependent on the degree to which the dimer subunit has been perturbed.

All three mutant *Pf*LDH structures clearly show electron density for a glycerol molecule at the dimer interface of the tetramer. This is interesting because 10% glycerol is known to stabilise the tetramer.Thus, it may be possible to further enhance the stabilisation of the oligomeric state of the enzyme by using other small solutes such as sucrose and glucose.

Recombinant Triple Mutant Pig M_4 LDH

Pig M_4 L-LDH is used as a model for human M_4 L-LDH in antimalarial selectivity studies. In order to overproduce the enzyme, the gene encoding the pig muscle LDH had previously been ligated into a pKK plasmid, this was then transfomed into an *E.coli* strain. On sequencing of the gene 3 sets of non-silent mutations were discovered. This had the effect of mutating 3 residues in the protein. Even though the mutations are at remote sites from the active site, the recombinant triple mutant pig M_4 LDH has a k_{cat} drop of 40% compared to the wild type. The crystal structure of the mutant has been solved. The model shows that the loss in activity of the mutant could probably be attributed to a destabilising change due to the S202N mutation.

Acknowledgements

I would like to acknowledge my supervisors Prof. J. John Holbrook and Dr. R. Leo Brady for their support during the research undertaken in this thesis. I would also like to thank Mark B. and John B. for attempting to answer my innumerable questions; all the other members of the 'crystallography corridor'; and also Kath Moreton for her helpful discussions on protein purification.

(PS. 'Other members of the 'crystallography corridor' include: Ailsa, Becky, Camille, Jon R., Julie, Kay, Mark T. and Vicky !)

Declaration

The work described in this thesis was conducted between October 1996 and August 2000 and unless otherwise acknowledged is that of the author.

B. S. Dhaliwal

List of Abbreviations

B.sterothermophilus (bs)	*Bacillus sterothermophilus*
E.coli	*Escherichia coli*
H.methylovorum	*Hyphomicrobium methylovorum*
L.casei	*Lactobacillus casei*
L.bulgaricus	*Lactobacillus delbrueckii* subsp. *bulgaricus*
L.helveticus	*Lactobacillus helveticus*
P.falciparum (*Pf*)	*Plasmodium falciparum*
ADH	alcohol dehydrogenase
D-GDH	glycerate dehydrogenase
D-HicDH	D-hydroxyisocaproate dehydrogenase or Broad specificity D-hydroxyacid dehydrogenase
D-LDH	D- isomer specific lactate dehydrogenase
D-3PGDH	D- 3-phosphoglycerate dehydrogenase
HADH	Short-chain L-3-hydroxacyl-CoA dehydrogenase
L-HicDH	L-hydroxyisocaproate dehydrogenase or Broad specificity L-hydroxyacid dehydrogenase
L-LDH	L- isomer specific lactate dehydrogenase
LDH	lactate dehydrogenase
MDH	malate dehydrogenase
ADP	Adenosine diphosphate
AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
cAMP	cyclic adenosine 3',5'-monophosphate
CRP	cAMP receptor protein
DDT	dichlorodiphenyl-trichloroethane
DHAP	Dihydroxyacetone phosphate
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylene diamine tetraacetic acid
GNDA	Gossylic nitrile diacetate
HEPES	*N*- 2- hydroxyethylpiperazine- *N'*-2- ethane sulphonic acid
IPTG	Isopropyl β-D-thiogalactoside
L Broth	Luria broth
LB Agar	Luria-Bertani agar
MES	2- [N-Morpholino] ethane sulphonic acid
NAD^+	Nicotinamide adenine dinucleotide (oxidized form)
NADH	Nicotinamide adenine dinucleotide (reduced form)
$NADP^+$	Nicotinamide adenine dinucleotide phosphate (oxidized form)
NADPH	Nicotinamide adenine dinucleotide phosphate (reduced form)
PAGE	Polyacrylamide gel electrophoresis
PEG	Polyethylene glycol
RNA	Ribonucleic acid

SDS	Sodium dodecyl sulphate
Se-Met	Selenomethionine
TEA	Triethanolamine
tNADH	1,4,5,6- tetrahydronicotinamide adenine dinucleotide
UV	Ultra-violet radiation
YT	Yeast tryptone
A	Absorbance
°C	Degrees Celsius
D	Daltons
g	Relative centrifugal force (in multiples of gravitational force)
g	Gram
pI	Isoelectric point
K	Kelvin
l	Litre
m	Metre
min	Minute
mol	mole
M	Molar
O.D.	Optical density
rpm	Revolutions per minute
U	Unit of enzyme activity (µmol NADH oxidized / min)
Å	Angstom units ($1x10^{-10}$ metres)
a, b, c, α, β, γ	Real space unit cell dimensions
B-factor	Temperature factor
d_{min}	Minimum Bragg spacing
h, k, l	Reciprocal lattice points (and indices of a reflection)
F_{hkl}	Structure factor for the h, k, l reflection
$\|F_{obs}\|$	Observed structure factor amplitude
$\|F_{calc}\|$	Calculated structure factor amplitude
NCS	Non-crystallographic symmetry
MAD	Multiple Anomalous Dispersion or Multiwavelength Anomalous Diffraction
MIR	Multiple Isomorphous Replacement
MR	Molecular Replacement
PDB	Protein Data Bank
R_{cryst}	Crystallographic R-factor
R_{free}	Free R-factor
r.m.s.d.	Root-mean-square deviation
SIR	Single Isomorphous Replacement
SRS	Synchrotron radiation source
x, y, z	Real space cartesian coordinates
MS	Mass Spectrometry
MALDI	Matrix-Assisted Laser Desorption Ionization
ESI	Electrospray Ionization

Unit Prefix

k	Kilo (10^3)
c	Centi (10^{-2})
m	Milli (10^{-3})
μ	Micro (10^{-6})
n	Nano (10^{-9})
p	Pico (10^{-12})
f	Femto (10^{-15})

Nucleotides

A	Adenine
T	Thymine
C	Cytosine
G	Guanine

Amino Acids

One letter code	Three letter code	Full name
A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartate
E	Glu	Glutamate
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine

Amino Acid Mutation Nomenclature

e.g. H205Q: The 205th amino acid residue histidine has been replaced by a glutamine residue.

Table of Contents

Abstract		i
Acknowledgements		ii
Declaration		iii
Abbreviations		iv
Table of Contents		vii
Chapter 1: L and D -2-Hydroxyacid Dehydrogenases		
1.1	Enzymes	1
1.2	Dehydrogenases	2
1.3	NAD^+-dependent 2-Hydroxyacid Dehydrogenases	3
1.3.1	The NAD^+ and $NADP^+$ Coenzymes	3
1.3.2	The NAD^+-dependent 2-Hydroxyacid Dehydrogenase Reaction	4
1.3.3	Stereospecificity of NAD^+-dependent 2-Hydroxyacid Dehydrogenases	5
1.3.3.1	The Hydride Transfer to NAD^+ is Stereospecific	5
1.3.3.2	Substrate Stereospecificity	6
1.4	NAD^+-dependent L-2-Hydroxyacid Dehydrogenases	7
1.4.1	Amino Acid Sequence Comparison	7
1.4.2	The General Structure of NAD^+-dependent L-2-Hydroxyacid Dehydrogenases	9
1.4.2.1	Dehydrogenases have Modular Polypeptide Chains	9
1.4.2.2	The NAD^+-binding Domain	9
1.4.2.3	NAD^+ / $NADP^+$ Discrimination	12
1.4.2.4	The Substrate-binding Domain	13

1.4.2.5	The Active Site	13
1.4.2.6	The Active Site Loop and its Role in Substrate Specificity	14
1.4.3	General Reaction Mechanism of NAD^+-dependent L-2-hydroxyacid Dehydrogenases	16
1.4.4	Examples of L-2-Hydroxyacid Dehydrogenases	20
1.4.4.1	L-Lactate Dehydrogenase	20
1.4.4.2	L-Malate Dehydrogenase	22
1.4.4.3	L-2-Hydroxyisocaproate Dehydrogenase	23
1.5	NAD^+-dependent D-2-Hydroxyacid Dehydrogenases	25
1.5.1	A Distinct Family of 2-Hydroxyacid Dehydrogenases	25
1.5.2	Amino Acid Sequence Comparison	26
1.5.3	Members of the D-Hydroxyacid Dehydrogenase Family	28
1.5.3.1	*VanH* gene product	28
1.5.3.2	D-Lactate Dehydrogenase	28
1.5.3.3	D-Glycerate Dehydrogenase	33
1.5.3.4	D-3-Phosphoglycerate Dehydrogenase	33
1.5.3.5	Erythronate-4-Phosphate Dehydrogenase	34
1.5.3.6	D-2-Hydroxyisocaproate Dehydrogenase	35

Chapter 2: Crystallographic Techniques

2.1	Protein Crystallisation	36
2.1.1	Introduction	36
2.1.2	Principles of Protein Crystallisation	37
2.1.3	Crystallisation Techniques	38
2.1.4	Cryo-cooling Crystals	39
2.2	X-ray Sources and Detectors	40

2.2.1	Introduction	40
2.2.2	X-ray Sources	40
2.2.2.1	Rotating anode generators	41
2.2.2.2	Synchrotron radiation	43
2.2.3	X-ray Detectors	44
2.2.3.1	Introduction	44
2.2.3.2	Image plates	45
2.3	X-ray Diffraction Theory	46
2.3.1	The Crystal	46
2.3.2	X-ray Diffraction	47
2.4	Molecular Structure from Diffraction Data	50
2.4.1	Calculating Electron Density	50
2.4.2	The Phase Problem	52
2.4.3	The Molecular Replacement Method	54
2.4.3.1	Introduction	54
2.4.3.2	The Rotation Function	55
2.4.3.3	The Translation Function	57
2.4.3.4	Rigid body refinement	57
2.4.4	The Multiple Isomorphous Replacement Method	57
2.4.4.1	Introduction	57
2.4.4.2	Obtaining Phases from Derivative Data	58
2.4.4.3	Locating heavy atom positions	63
2.4.4.4	Preparing heavy atom derivative crystals	65
2.4.5	The Multiwavelength Anomalous Dispersion Method	66
2.4.5.1	Introduction	66

2.4.5.2 Anomalous Scattering 67

2.4.5.3 Effects of anomalous scattering 68

2.4.5.4 Obtaining Phases from anomalous scattering data 70

2.4.5.5 Locating anomalous scatters 73

2.4.5.6 Choosing wavelengths for MAD data collection 74

2.4.5.7 Incorporation of anomalous scatters 75

2.5 Protein Structure Refinement 77

2.5.1 The Initial Model 77

2.5.2 Structure Refinement 77

2.5.2.1 Introduction 77

2.5.2.2 Refinement of protein structures 78

2.5.2.3 Maximum likelihood refinement 78

2.5.2.4 Least squares refinement 78

2.5.2.5 Refinement by simulated annealing 79

2.5.2.6 Rebuilding the model 80

2.5.2.7 Monitoring the progress of refinement 80

Chapter 3: *Lactobacillus delbruekii* subsp. *bulgaricus* D-2-Hydroxyisocaproate Dehydrogenase: A Broad Specificity D-2-Hydroxyacid Dehydrogenase

3.1 D-2-Hydroxyisocaproate Dehydrogenase 81

3.2 The H205Q Mutation 86

Chapter 4: Purification, Crystallisation and Diffraction Data Collection of the Wild Type & H205Q mutant D-2-Hydroxyacid Dehydrogenase

4.1	Purification of D-HicDH	90
4.2	Protein Concentration	94
4.3	Crystallisation and Diffraction Data Collection of the Wild Type & H205Q mutant D-2-Hydroxyacid Dehydrogenase	96
4.3.1	Crystallisation Methods	96
4.3.2	Data Processing and Reduction	96
4.3.3	Crystallising the wild type & H205Q apoenzyme	97
4.3.4	Crystallising the wild type & H205Q binary complexes	101
4.3.5	Crystallising the wild type & H205Q ternary complexes	110
4.3.5.1	1,4,5,6-Tetrahydronicotinamide Adenine Dinucleotide	110
4.3.5.2	Crystallising the abortive ternary complex	110

Chapter 5: Attempts at Phasing the D-2-Hydroxyisocaproate Dehydrogenase Diffraction Datasets

5.1	Molecular Replacement	115
5.1.1	Choosing the Search Models	115
5.1.2	Estimating the Number of Molecules in the Asymmetric Unit	116
5.1.3	Attempts at solving the D-HicDH structure using *AMoRe*	117
5.1.3.1	The *AMoRe* program	117
5.1.3.2	Sorting and Tabling	117
5.1.3.3	The Rotation Function	119

5.1.3.4	The Translation Function and Fitting	119
5.1.3.5	Attempts at Phasing the D-HicDH datasets	119
5.1.4	Attempts at Solving the D-HicDH Structure using AMoRe with Normalised Structure Factor Amplitudes	126
5.1.4.1	Normalised Structure Factor Amplitudes	126
5.1.4.2	Calculating *E* values, Sorting & Tabling, and the Rotation Function	127
5.1.4.3	The Translation Function using *E* values	127
5.1.4.4	Attempts at Phasing the D-HicDH datasets using *E*'s	127
5.1.5	Attempts at solving the D-HicDH structure using *EPMR*	129
5.1.6	Conclusions	131
5.2	Multiple Isomorphous Replacement	132
5.2.1	Heavy atom soaking procedures	132
5.2.2	The Heavy atom salts utilised	133
5.2.3	Diffraction Data Collection of Potential Derivatives	134
5.2.4	Location and Refinement of possible Heavy Atom Sites	135
5.2.4.1	Scaling the derivative and native datasets	136
5.2.4.2	Locating the heavy atom sites	136
5.2.4.3	Refinement of heavy atom parameters	137
5.2.4.4	Locating and refining possible heavy atoms sites in the H205Q 'derivatives'	137
5.2.5	Conclusions	139
5.3	Multiwavelength Anomalous Dispersion	140
5.3.1	Introduction	140
5.3.2	Preparation of selenomethionyl D-HicDH protein for phase determination	140
5.3..2.1	Methionine biosynthetic pathway inhibition	141

5.3.2.1.1	Inhibiting aspartokinases	141
5.3.2.1.2	Bacterial growth in selenomethionine containing media	141
5.3.2.1.3	Preliminary investigations into AR58p[GIN113/6] bacterial cell growth on selenomethionyl rich media	142
5.3.2.1.4	Se-Met D-HicDH purification protocol using the AR58p[GIN113/6] expression system	145
5.3.2.1.5	Using JM105p[KK183] cells and catabolite repression to reduce pre-induction D-HicDH expression	148
5.3.2.1.6	Se-Met D-HicDH purification protocol using the JM105p[KK183] expression system	151
5.3.3	Mass Spectrometry	154
5.3.3.1	Electrospray Ionization Mass Spectrometry	154
5.3.3.2	Matrix-Assisted Laser Desorption Ionization Mass Spectrometry	155
5.3.3.3	Mass Spectrometric analysis of selenomethionyl D-HicDH protein produced by inhibiting methionine biosynthesis	156
5.3.4	Selenomethionyl protein expression using a methionine auxotroph	161
5.3.4.1	The DL41 and DL41(DE3) Methionine Auxotrophic E.coli cell strains	161
5.3.4.2	pKK 183 plasmid DNA extraction	163
5.3.4.3	Preparation of Competent DL41(DE3) cells	164
5.3.4.5	Transformation of the pKK 183 plasmid into the competent DL41(DE3) cell strain	165
5.3.4.6	Prepararion of DL41(DE3)p[KK183] glycerol stocks	165
5.3.4.7	Assesing the transformation	165
5.3.4.8	Se-Met D-HicDH purification protocol using the DL41(DE3)p[KK183] methionine auxotrophic expession system	169
5.3.4.9	Mass spectrometric analysis of selenomethionyl D-HicDH protein produced by using a methionine auxotroph	170

5.3.4.10 Crystallising the selenomethionyl H205Q D-HicDH protein produced using a methionine auxotroph 171

5.3.5 Attempts at MAD phasing 172

5.3.5.1 X-ray Fluoresence: Choosing wavelengths for MAD data collection 172

5.3.5.2 Attempts at data collection at the absorption edge 173

5.3.6 Conclusions 173

Chapter 6: Subunit Interface Mutants of L-Lactate Dehydrogenase from the Malarial Parasite *Plasmodium falciparum*

6.1 Malaria 174

6.1.1 Introduction 174

6.1.2 The Causes of Malaria 174

6.1.3 Combating Malaria 175

6.2 *Plasmodium falciparum* L-Lactate Dehydrogenase 177

6.3 Mutations at the Subunit Interface of *Pf*LDH 179

Chapter 7: Determination and Analysis of the Molecular Structures of the C183 *Pf*LDH Subunit Interface Mutants

7.1 Aims 182

7.2 Crystallisation of the Subunit Interface Mutant *Pf*LDH Ternary Complexes 182

7.3 Data Collection 183

7.4 Solving the C183 Mutant *Pf*LDH Structures 187

7.4.1 Introduction 187

7.4.2 Solving the C183V Mutant 188

7.4.3	Solving the C183AMutant	189
7.4.4	Solving the C183SMutant	190
7.5	Crystallographic Refinement and Model Building of the C183 Mutant *Pf*LDH Structures	191
7.6	Validation of the C183 Mutant *Pf*LDH Models	203
7.7	Analysis of the C183 Mutant *Pf*LDH Molecular Structures	213
7.7.1	Reduced Enzymatic Rates of the C183 *Pf*LDH Mutants	213
7.7.2	Oxidative Stability of the C183 *Pf*LDH Mutants	215
7.8	Conclusions	218

Chapter 8: Crystallisation and Preliminary Characterisation of a Recombinant Triple Mutant Pig M_4 L-LDH

8.1	Introduction	220
8.1.1	Mammalian LDH Isoenzymes	220
8.1.2	A Recombinant Pig Triple Mutant Pig M_4 LDH	221
8.2	Protein Purification and Crystallisation	222
8.2.1	Protein Expression	222
8.2.2	Protein Purification	223
8.2.3	Crystallisation	225
8.3	Data Collection	226
8.4	Solving the Pig M_4 LDH Triple Mutant Structure	228
8.5	Crystallographic Refiment and Model Building	232
8.6	Discussion	238
Summary		240
References		243

Chapter 1

L and D -2-Hydroxyacid Dehydrogenases

1.1 Enzymes

The enormous variety of biochemical reactions that comprise life are nearly all mediated by a series of biological catalysts known as enzymes.

Enzymes differ from chemical catalysts in several respects. Firstly, the rates of enzyme catalysed reactions are at least several orders of magnitude greater than those of the corresponding chemically catalysed reactions, and typically 10^6 to 10^{12} orders of magnitude greater than the uncatalysed reactions. Secondly, the catalytic activities of many enzymes are regulated; this control can be mediated through the concentration of the enzyme's substrates, products or some other molecule. The regulatory process initiated could lead to a covalent or allosteric modification of the catalyst, or a change in the amounts of enzyme synthesised. Furthermore, in contrast to chemical catalysts, enzyme catalysed reactions occur in milder conditions. For instance, most biochemical reactions occur below 100°C (most occuring between 25-40°C), at around neutral pH, and usually at atmospheric pressure. Another advantage of enzymes over conventional chemical catalysts is that the biochemical catalysts are highly specific: An enzyme usually catalyses a single chemical reaction or a set of closely related reactions.

These properties make enzymes extremely useful as catalysts for a broad range of organic synthetic reactions used in the pharmaceutical industries, with their capacities for producing chirally pure compounds being the most exploited.

A class of enzymes widely applied in organic synthesis is the dehydrogenases.

1.2 Dehydrogenases

Physiologically, the role of the dehydrogenases is to catalyse important redox reactions involved in the metabolism and energy conversion of living cells.

There are many dehydrogenase subfamilies; these include the short-chain, medium-chain and 2-hydroxyacid dehydrogenases.

The short-chain dehydrogenase subfamily consists of enzymes such as short-chain L-3-hydroxyacyl-coA dehydrogenases (HADH) which catalyse the third step in the fatty acid β-oxidation pathway in mitochondria, converting L-3-hydydroxyacyl-Coenzyme A to 3-ketoacyl-coenzyme with the concurrent reduction of NAD^+ (Noyes and Bradshaw, 1973).

The medium-chain dehydrogenase subfamily is comprised of enzymes such as the various classes of medium-chain alcohol dehydrogenases (ADH) that cataylse the oxidation of alcohols to aldehydes using NADH as a cofactor (Schwartz *et al.*, 1979).

From an industrial point of view, one of the most valuable groups of dehydrogenases are the NAD^+- (and to a lesser extent $NADP^+$-) -dependent-2-hydroxyacid dehydrogenases (Hummel and Kula, 1989). This is because chirally pure 2-hydroxyacids are valuable compounds for the production of various pharmaceuticals such as penicillins, cephalosporins, anti-hypertensives, anti-obesity compounds, sweetners, etc. Also, the 2-hydroxyacid dehydrogenases themselves are often the targets of inhibitory rational drug-design (*e.g.* Dunn *et al.*, 1996).

1.3 NAD^+-dependent 2-Hydroxyacid Dehydrogenases

Despite the widespread use of 2-hydroxyacid dehydrogenases, much remains to be learned about the factors that determine the specificity and stereospecificity of these enzymes. However, because of the increasingly broad spectrum of new or unnatural substrates that the enzymes are called upon to catalyse, it is becoming essential that the structural basis of substrate specificity and stereospecificity is understood. As well as advancing knowledge of the physiological functions performed by the enzymes, this information would also aid in the rational modification of enzyme specificity through site-directed mutagenesis.

1.3.1 The NAD^+ and $NADP^+$ Coenzymes

The coenzyme nicotinamide adenine dinucleotide (NAD^+) is a major electron acceptor in the oxidation of molecules. The reactive part of NAD^+ is its nicotinamide ring. In the oxidation of a substrate the nicotinamide ring accepts a hydrogen ion and two electrons (equivalent to a hydride ion). The reduced form of the electron carrier is known as NADH (Figure 1.1).

The electron donor in most reductive biosynthetic processes is NADPH, the reduced form of nicotinamide adenine dinucleotide phosphate ($NADP^+$). NADPH differs from NADH in that the 2′-hydroxyl group of its adenosine moiety is esterified with phosphate.

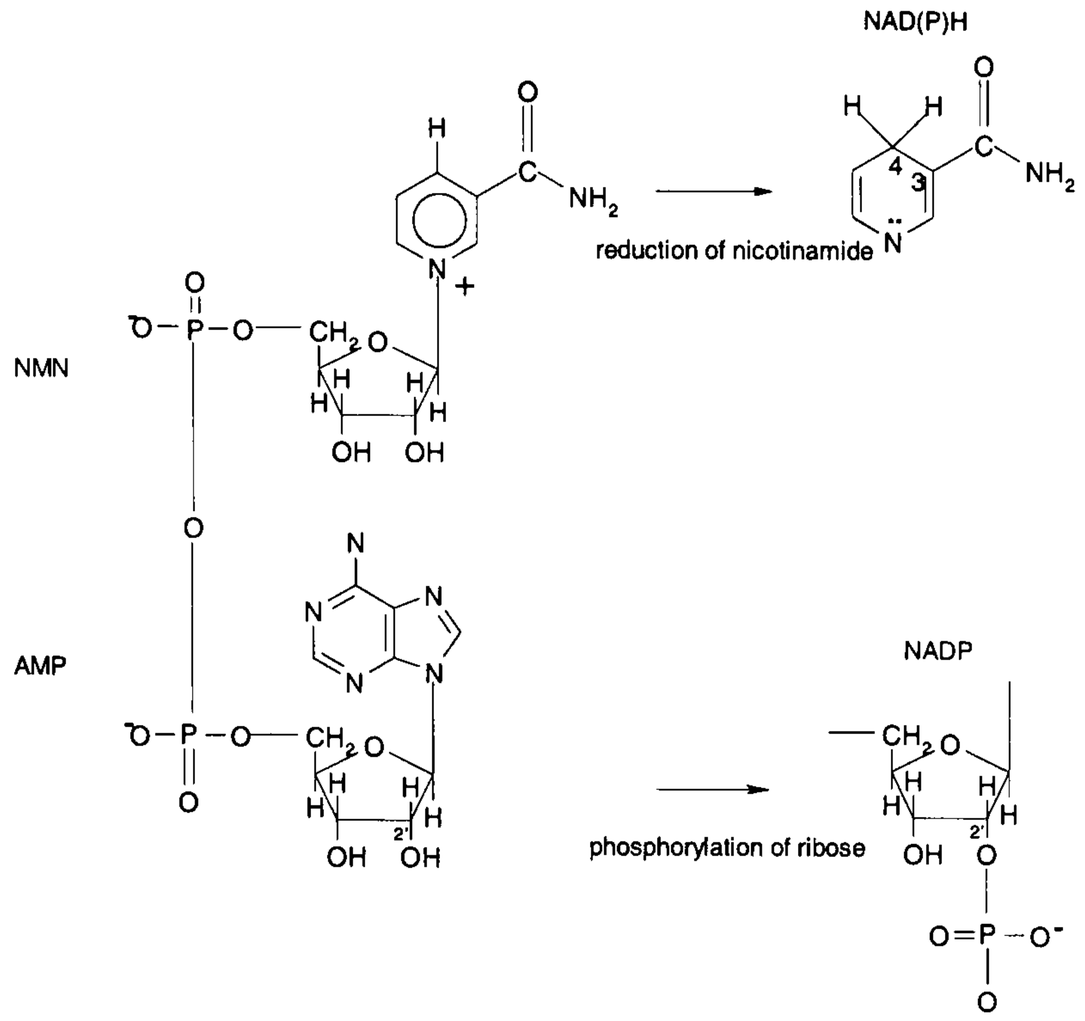


Figure 1.1 The structures of NAD^+ and $NADP^+$. The cofactors bind to dehydrogenases in an energetically unfavourable form, shaped like a boomerang, that is stabilised by interactions with the protein (Holbrook *et al.*, 1975).

1.3.2 The NAD^+-Dependent 2-Hydroxyacid Dehydrogenase Reaction

NAD^+-dependent 2-hydroxyacid dehydrogenases catalyse the stereospecific and reversible reduction of aliphatic and aromatic 2-oxocarboxylic acids to the corresponding 2-hydroxycarboxylic acids (Figure 1.2).

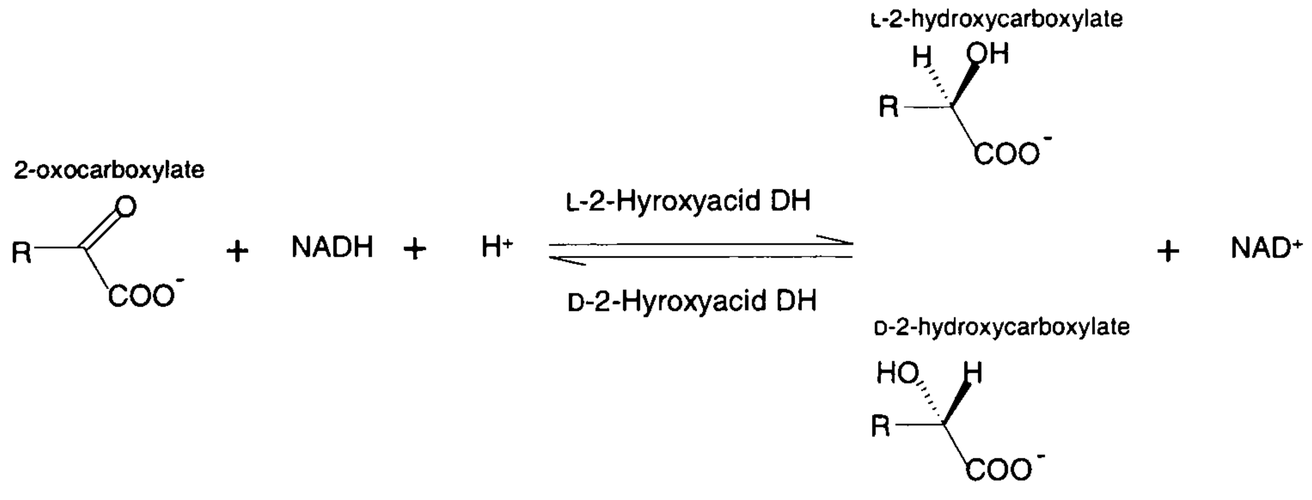


Figure 1.2 General reaction of NAD^+-dependent 2-Hydroxyacid Dehydrogenases. (R: substrate side-chain).

The enzymes distinguish between the stereospecific isomers of their substrates and bind only one of them. Therefore on this basis the NAD^+-dependent 2-hydroxyacid dehydrogenases can be divided into two groups depending on

whether they bind the D or L enatiomer of the 2-hydroxyacid substrate (Clarke and Daffron, 1998).

1.3.3 Stereospecificity of NAD^+-dependent 2-Hydroxyacid Dehydrogenases

1.3.3.1 The Hydride Transfer to NAD^+ is Stereospecific

Oxidation of the the hydroxyacid involves the removal of two hydrogen atoms. One is transferred as a hydride ion H^- to the dihydronicotinamide moiety of the coenzyme, the other is releasesd as a proton (Fersht, 1985).

In dehydrogenases the hydride transfer to NAD^+ is stereospecific : Using a deuterated substrate, so that the enzyme tranferred a deuterium atom to NAD^+ instead of a hydrogen atom , it was found that some dehydrogenases transferred the hydride to one side of the ring, and others to the opposite side (Weinhold *et al.*, 1991) (Figure 1.3).

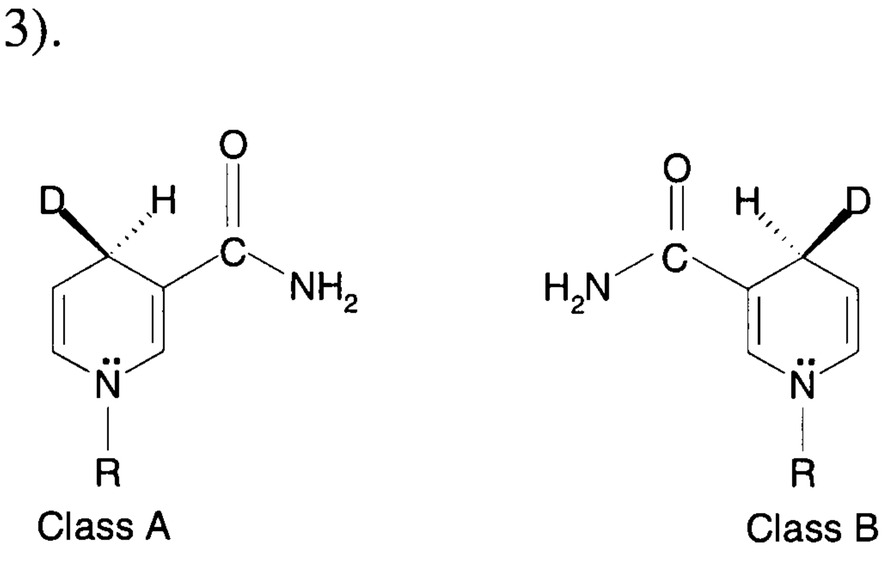


Figure 1.3 Stereospecific hydride transfer to NAD^+. (D: deutrium atom).

Therefore, for many years the NAD^+-dependent dehydrogenases have been further divided into two groups depending on the stereospecificity of hydrogen transfer to the C4 atom of the dihydronicotinamide moiety. The C4 position of NADH is a prochiral centre (an atom of a molecule which becomes a chirality centre by replacing one of the two stereoheterotopic ligands attached to it by a different ligand is said to be a prochirality centre *e.g.* C-1 of ethanol; C-3 of butan-2-ol). Class A NAD^+-dependent dehydrogenases transfer a hydride ion to

the *pro-R* (A-side) face of NAD^+, whereas the class B enzymes transfer to the *pro-S* (B-side) face.

There is no metabolic or mechanistic reason for the existence of the two classes of dehydrogenase: The difference in classes A and B is due only to differences in the structures of the catalytic domains. The orientation of the bound nicotinamide group is determined by specific hydrogen bonds to the amide group in one orientation and steric clashes between amide and protein groups in the other. The coenzyme conformations in the two classes of enzymes are similar except that the orientations of the nicotinamide rings differ by ~180° (Holbrook *et al.*, 1975). For a specific enzyme in different species the stereochemistry is maintained because the conformation of the enzyme remains preserved.

1.3.3.2 Substrate Stereospecificity

Another aspect of stereospecificity in NAD^+-dependent 2-hydroxyacid dehydrogenase reactions concerns the substrate: Only one of the two possible enatiomers of the 2-hydroxyacid is formed on reducing the 2-ketoacid substrate. The achiral ketoacid is bound and presented to the enzyme in such a way that only one side of the substrate is ready to accept transfer of the hydride ion from C4 of dihydronicotinamide. The same catalytic mechanism can be used to reduce a ketoacid to an L- or D-hydroxyacid (*cf.* hydride transfer in L- and D-LDHs, below).

The NAD^+-dependent 2-hydroxyacid dehydrogenases' chiral selectivity is extreme, for instance, for L-lactate dehydrogenase the chirality rule is only broken once in 10^8 turnovers (Anderson and La Reau, 1988).

This property of enatioselectivity makes the group of enzymes attractive as catalysts for the pharmaceutical industry.

1.4 NAD^+-dependent L-2-Hydroxyacid Dehydrogenases

1.4.1 Amino Acid Sequence Comparison

The NAD^+-dependent L-2-hydroxyacid dehydrogenases comprise one of the largest and best-studied families of nucleotide-binding proteins. Most members of the family share a common structure-function relationship. The amino acid sequence comparison of three members of the enzyme family from various sources are shown in Figure 1.4.

The sequence alignment shows that the members of the NAD^+-dependent L-2-hydroxyacid dehydrogenase family are highly homologous, and that the pig L-LDH isoenzymes are virtually identical to the human L-LDH isoenzymes (see Chapter 8). The catalytic residues are also shown to be conserved in all proteins, these residues include; R109 (in the L-LDH numbering system, based upon the sequence of dogfish muscle L-LDH (Rossmann *et al.*, 1971)), D168, R171 and H195.

```
                1                                                                                                   99
DogfishL-LDH-M      ATLKDKLIGH LATSQEPRSY NKITVVGVGA VGMACAISIL MKDLADEVAL VDVMEDKLKG EMMDLQHGSL F-LHTAKIVS GKDYSVSAGS KLVVITAGAR
PigL-LDH-M          ATLKDQLIHN LL-KEEHVPH NKITVVGVGA VGMACAISIL MKELADEIAL VDVMEDKLKG EMMDLQHGSL F-LRTPKIVS GKDYNVTANS RLVVITAGAR
HumanL-LDH-M        ATLKDQLIYN LL-KEEQTPQ NKITVVGVGA VGMACAISIL MKDLADELAL VDVIEDKLKG EMMDLQHGSL F-LRTPKIVS GKDYNVTANS KLVIITAGAR
PigL-LDH-H          ATLKEKLIAP VAEEETTIPN NKITVVGVGQ VGMACAISIL GKSLTDELAL VDVLEDKLKG EMMDLQHGSL F-LQTPKIVA DKDYSVTANS KIVVVTAGVR
HumanL-LDH-H        ATLKEKLIAP VAEEEATVPN NKITVVGVGQ VGMACAISIL GKSLADELAL VDVLEDKLKG EMMDLQHGSL F-LQTPKIVA DKDYSVTANS KIVVVTAGVR
HumanL-LDH-X        STVKEQLIEK LIEDDEN-SQ CKITIVGTGA VGMACAISIL LKDLADELAL VDVALDKLKG EMMDLQHGSL F-FSTSKVTS GKDYSVSANS RIVIVTAGAR
P.falciparumL-LDH                    MAPK AKIVLVGSGM IGGVMATLIV QKNLGD-VVL FDIVKNMPHG KALDTSHTNV MAYSNCKVSG SNTYDDLAGA DVVIVTAGFT
L.confususL-HicDH                       A RKIGIIGLGN VGAAVAHGLI AQGVADDYVF IDANEAKVKA DQIDFQDAMA NLEAHGNIVI -NDWAALADA DVVISTLGNI
HumanL-MDH-A      SEPIRVL VTGAAGQIAY SLLYSIGNGS VFGKDQPIIL V------LLD ITPMMGVLDG VLMELQDCAL PLLKDV-IAT DKEDVAFKDL DVAILVGSMP
 Consensus  .........       . .......... .ki..!G.G. !g.a.a..il .k.lad...l .d....klkg ..m#lqh..l ..l...k!.. .kdy...a.. dvv!.tag..

                100                                                                                                 194
DogfishL-LDH-M      QQEGES---- -RLNLVQRNV NIFKFIIPDI VKHSPDCI-I LVVSNPVDVL TYVAWKLSGL PMHRIIGSGC NLDSARFRYL MGERLGVHSS SCHGWVI-GE
PigL-LDH-M          QQEGES---- -RLNLVQRNV NIFKFIIPNI VKYSPNCK-L LVVSNPVDIL TYVAWKISGF PKNRVIGSGC NLDSARFRYL MGERLGVHPL SCHGWIL-GE
HumanL-LDH-M        QQEGES---- -RLNLVQRNV NIFKFIIPNV VKYSPNCK-L LIVSNPVDIL TYVAWKISGF PKNRVIGSGC NLDSARFRYL MGERLGVHPL SCHGWVL-GE
PigL-LDH-H          QQEGES---- -RLNLVQRNV NVFKFIIPQI VKYSPDCI-I IVVSNPVDIL TYVTWKLSGL PKHRVIGSGC NLDSARFRYL MAEKLGVHPS SCHGWIL-GE
HumanL-LDH-H        QQEGES---- -RLNLVQRNV NVFKFIIPQI VKYSPDCI-I IVVSNPVDIL TYVTWKLSGL PKHRVIGSGC NLDSARFRYL MAEKLGIHPS SCHGWIL-GE
HumanL-LDH-X        QQEGET---- -RLALVQRNV AIMKSIIPAI VHYSPDCK-I LVVSNPVDIL TYIVWKISGL PVTRVIGSGC NLDSARFRYL IGEKLGVHPT SCHGWII-GE
P.falciparumL-LDH   KAPGKSDKEW NRDDLLPLNN KIMIEIGGHI KKNCPNAF-I IVVTNPVDVM VQLLHQHSGV PKNKIIGLGG VLDTSRLKYY ISQKLNVCPR DVNAHIV-GA
L.confususL-HicDH   KLQQDNPTG- DRFAELKFTS SMVQSVGTNL KESGFHGV-L VVISNPVDVI TALFQHVTGF PAHKVIGTGT LLDTARMQRA VGEAFDLDPR SVSGYNL-GE
HumanL-MDH-A        RREGME---- -RKDLLKANV KIFKSQGAAL DKYAKKSVKV IVVGNPANTN CLTASKSAPS IPKENFSCLT RLDHNRAKAQ IALKLGVTAN DVKNVIIWGN
 Consensus          ..eg..... .R..ll..nv .ifksig..i .ky.p..... iV!sNPv#.. t....k.sg. p...vig.g. .LD.aR..y. i.eklgv.p. sv.g.i..Ge

                195                                                                                                 291
DogfishL-LDH-M      HGDSSVPVWS GMNVAGVSLK E-LHPELGTD KDKENWKKLH KDVVDSAYEV IKLKGYTSWA IGLSVADLAE TIMKNLCRVH PVSTMV---K DFYGIKNDVF
PigL-LDH-M          HGDSSVPVWS GVNVAGVSLK N-LHPELGTD ADKEHWKAVH KQVVDSAYEV IKLKGYTSWA IGLSVADLAE SIMKNLRRVH PISTMI---K GLYGIKEDVF
HumanL-LDH-M        HGDSSVPVWS GMNVAGVSLK T-LHPDLGTD KDKEQWKEVH KQVVESAYEV IKLKGYTSWA IGLSVADLAE SIMKNLRRVH PVSTMI---K GLYGIKDDVF
PigL-LDH-H          HGDSSVAVWS GVNVAGVSLQ E-LNPEMGTD NDSENWKEVH KMVVESAYEV IKLKGYTNWA IGLSVADLIE SMLKNLSRIH PVSTMV---Q GMYGIENEVF
HumanL-LDH-H        HGDSSVAVWS GVNVAGVSLQ E-LNPEMGTD NDSENWKEVH KMVVESAYEV IKLKGYTNWA IGLSVADLIE SMLKNLSRIH PVSTMV---K GMYGIENEVF
HumanL-LDH-X        HGDSSVPLWS GVNVAGVALK T-LDPKLGTD SDKEHWKNIH KQVIQSAYEI IKLKGYTSWA IGLSVMDLVG SILKNLRRVH PVSTMV---K GLYGIKEELF
P.falciparumL-LDH   HGNKMVLLKR YITVGGIPLQ EFINNKLISD AELE---AIF DRTVNTALEI VNLHASPYVA PAAAIIEMAE SYLKDLKKVL ICSTLL---E GQYG-HSDIF
L.confususL-HicDH   HGNSQFVAWS TVRVMGQPIV T------LAD AGDIDLAAIE EEARKGGFTV LNGKGYTSYG VATSAIRIAK AVMADAHAEL VVSNRR---D DM-----GMY
HumanL-MDH-A        HSSTQYPDVN HAKVKLQGKE VGVYEALKDD SWLKG-EFVT TVQQRGAAVI KARKLSSAMS AAKAICDHVR DIWFGTPEGE FVSMGVISDG NSYGVPDDLL
 Consensus          Hg.s.vp.ws ...V.g..l. ......l..D ...e....!. ......a.e! ..lkgyt..a .a.s..d.a. .i.k.l.... .vSt...... ..yg...d.f

                292                                 336
DogfishL-LDH-M      LSLPCVLDNH GISNIVKMKL KPDEEQQLQK SATTLWDIQK DLKF
PigL-LDH-M          LSVPCILGQN GISDVVKVTL TPEEEAHLKK SADTLWGIQK ELQF
HumanL-LDH-M        LSVPCILGQN GISDLVKVTL TSEEEARLKK SADTLWGIQK ELQF
PigL-LDH-H          LSLPCVLNAR GLTSVINQKL KDDEVAQLKN SADTLWGIQK DLKDL
HumanL-LDH-H        LSLPCILNAR GLTSVINQKL KDDEVAQLKK SADTLWDIQK DLKDL
HumanL-LDH-X        LSIPCVLGRN GVSDVVKINL NSEEEALFKK SAETLWNIQK DLIF
P.falciparumL-LDH   GGTPVVLGAN GVEQVIELQL NSEEKAKFDE AIAETKRMKA LA
L.confususL-HicDH   LSYPAIIGRD GVLAETTLDL TTDEQEKLLQ SRDYIQQRFD EIVDTL
HumanL-MDH-A        YSFPVVIKNK TWKFVEGLPI NDFSREKMDL TAKELTEEKE SAFEFLSSA
 Consensus          ls.P.!lg.. g...v..l.l n..e..k... sa..l..... .........
```

Figure 1.4. Sequence alignment of NAD^+-dependent L-2-hydroxyacid dehydrogenases. M=Muscle isoenzyme, H=Heart isoenzyme, X=X isoenzyme. Residues in red are highly conserved (> 90%); residues in blue are not highly conserved (> 50%). If the remaining residues are conservative substitutions the consensus residue is in uppercase. Produced using *MultAlin* (Corpet,1988).

1.4.2 The General Structure of NAD^+-dependent L-2-Hydroxyacid Dehydrogenases

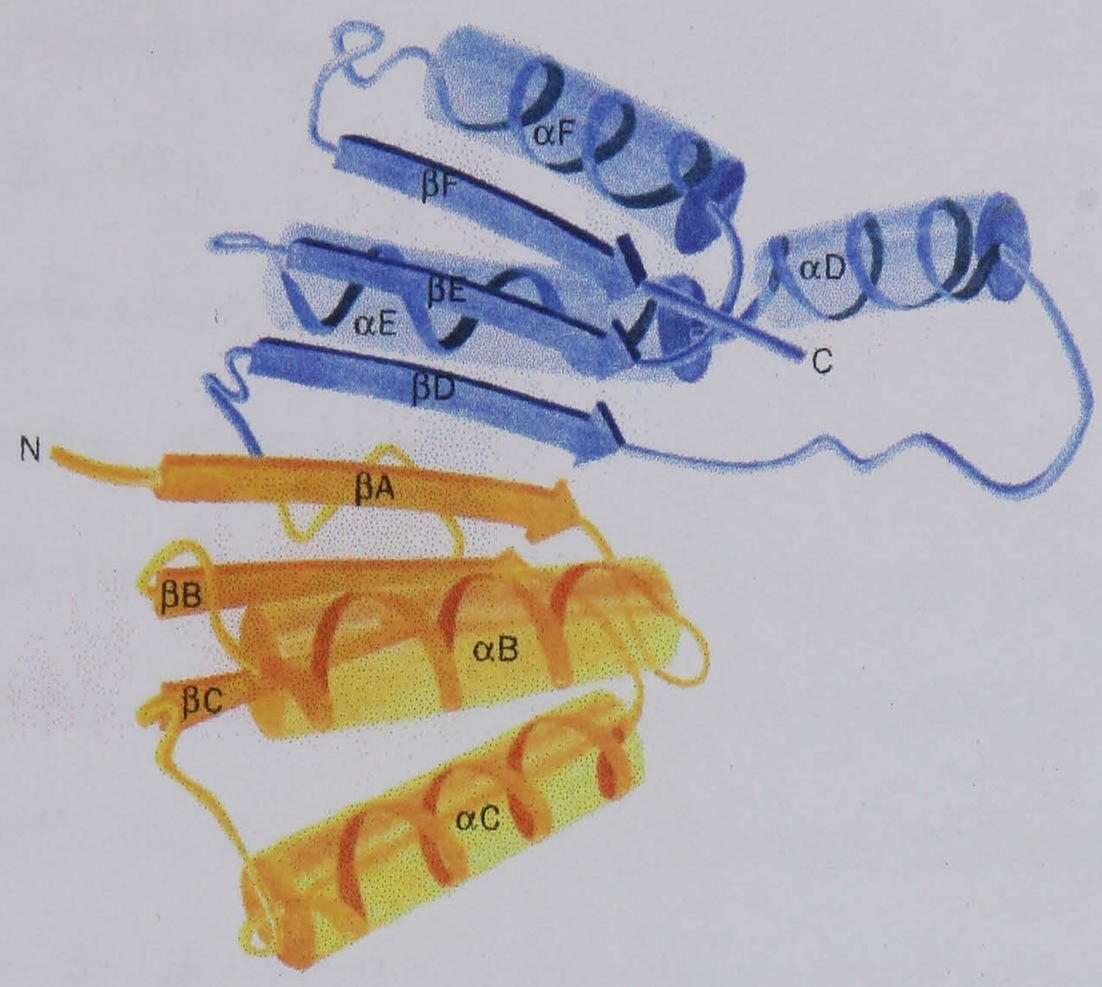
1.4.2.1 Dehydrogenases have Modular Polypeptide Chains

Dehydrogenases are in general fairly large proteins consisting of either a tetramer or a dimer of mostly identical polypeptide chains. The lengths of the chains vary slightly, but they are typically around 330 residues. The polypeptide chains fold into two distinct functional domains, known as the NAD^+- and substrate-binding domains. However, the functionally similar NAD^+-binding domains occur in different regions of the polypeptide in the dehydrogenases.

1.4.2.2 The NAD^+-binding Domain

Despite the lack of any amino acid sequence homology the NAD^+-binding domains of the 2-hydroxyacid dehydrogenases are structurally very similar (Figure 1.5).

(a)



(b)

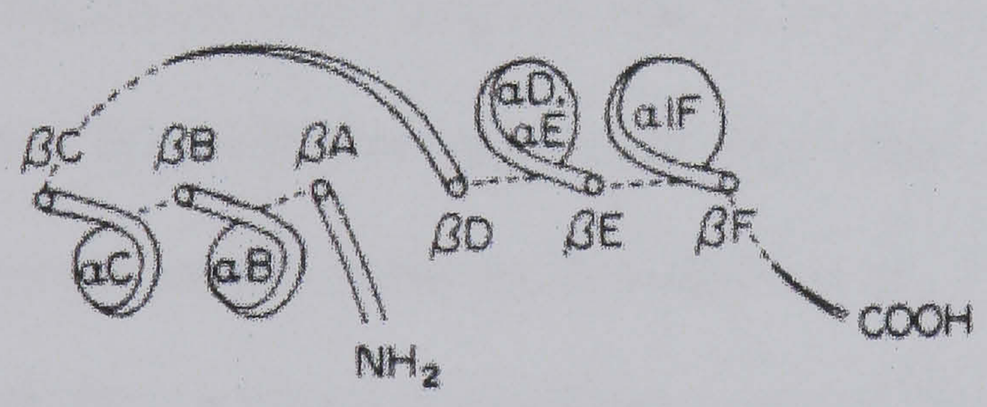


Figure 1.5 (a) An idealised representation of the NAD^+-binding domain from various L-2-hydroxyacid dehydrogenases. The yellow and blue segments each represent a mononucleotide binding domain. (b) Side-on view of the fold. (From Voet and Voet, 1995).

The dinucleotide-binding fold differs in detail from one dehydrogenase to another. However, in its idealised form the 80 - 100 amino acid fold consists of six strands of parallel β sheet with four parallel α helices running antiparallel to the sheet.

The NAD^+-binding domain consists of two approximately identical mononucleotide binding units related roughly by 2-fold symmetry (Rao and

Rossmann, 1973). 'The half-structure' is formed by a pair of β-α-β-α-β motifs: Each half of the domain binding one of the nucleotides in the dinucleotide NAD^+. The mononucleotide binding unit is also known to occur in other proteins, for example flavodoxin, which binds the mononucleotide FMN (Rossmann *et al.*, 1974).

The structurally conserved βαβαβ super-secondary motif of the dinucleotide-binding domain is commonly referred to as the Rossmann fold.

In most L-2-hydroxyacid dehydrogenases the regions of similar structure in the coenzyme-binding domains form the structural framework of the main chain that binds NAD^+. The actual interactions between the coenzyme and protein occur, however, through amino acid side chains, with the coenzyme binding in almost identical positions in the enzymes. But, NAD^+ binding involves different combinations of side chains in different enzymes. Indeed, sequence comparisons of the same enzyme from different species has shown that side chains vary as much as nonfunctional surface residues (Rossmann *et al.*, 1975). So, there are several ways of reaching a common structural goal, and this is reflected in the evolution of the enzymes. The absence of significant sequence homologies in the NAD^+-binding domain of the different dehydrogenases suggest that it is difficult to say anything about the evolutionary histories of these enzymes apart from that they appear to be very ancient.

Although most of the residues vary in the coenzyme-binding domain, there are still some conserved residues among dehydrogenases in this fold. This is due to the fact there are strong stereochemical constraints at specific positions in the fold that must be kept in order to maintain the structure and function of the

domain. These invariant chains make it possible to predict from amino acid sequence the regions in the polypeptide chain that are involved in coenzyme binding.

The βA-αB-βB-αC-βC motif (Figure 1.5) is a structurally conserved motif comprising of about 30 residues. It has been used to identify the NAD^+-binding domains where the dehydrogenase structures are unknown. This region contains three conserved glycine residues with the sequences Gly/Ala-X-Gly/Ala-X-X-Gly/Ala where X is any residue. The glycine-rich region is involved in a turn next to the adenine-ribose moiety (Rossmann *et al.*, 1975). Larger sidechains at these positions would disturb the structural framework of the fold and push NAD^+ out.

There are also six hydrophobic conserved residues which form a hydrophobic core between the α-helix and β-sheets.

1.4.2.3 NAD^+ / $NADP^+$ Discrimination

The NAD^+-binding domain also contains a conserved aspartate located some 25 residues from the glycine-rich region further towards the C-terminus (residue 53 in Figure 1.4). The residue's function is to discriminate between the coenzymes NAD^+ and $NADP^+$. NAD^+ binds to most L-2-hydroxyacid dehydrogenases through hydrogen bonds between its 2′-OH of the adenosine ribose moiety and the conserved aspartate. However, $NADP^+$ has a phosphate group attached to its 2′-OH and therefore it is prevented from binding to NAD^+-linked dehydrogenases by steric clashes and repulsions between the two negative groups. Instead of the aspartate residue, dehydrogenases that are dependent on $NADP^+$ have a small side chain, usually glycine, to make room for

the phosphate group and a positively charged side chain, such as arginine, nearby to bind the phosphate (Clarke and Daffron, 1998).

1.4.2.4 The Substrate-binding Domain

As well as recognising and binding a specific substrate enatiomer, the substrate-binding domain also provides the amino acid residues required for catalysis.

The substrate-binding domains of the 2-hydroxyacid dehydrogenases have completely different sequences as well as structures. In fact, apart from L-LDHs' and MDHs', a unique topology has been observed in each dehydrogenase.

1.4.2.5 The Active Site

The active site of the NAD^+-dependent dehydrogenases is located in a cleft between the NAD^+- and substrate-binding domains. The respective domains of the two binding sites are orientated so that the reactive part of the coenzyme, C4 of the nicotinamide ring, is in close proximity to the hydrogen atom to be transferred from the substrate.

The most important residues within the active site are arginine-171, histidine-195, aspartate-168 and threonine-246 (again, the numbering system is based upon dogfish muscle L-LDH). The histidine and aspartate residues are involved in proton transfer during catalysis. Threonine-246 hydrogen bonds with the substrate's carboxyl group, and this is complemented by a hydrogen bond between the substrate carbonyl and histidine-195. Both Thr246, His195 serve to orient the substrate through hydrogen bonding (Figure 1. 11, page 21).

Arginine171 also binds and orients the substrate in the active site by forming a bifurcated ion pair between its NH2 and NH1 groups and the carboxyl group of

the substrate. It is thought that this interaction may also play a small part in substrate specificity (Hart *et al.*, 1987).

1.4.2.6 The Active Site Loop and its Role in Substrate Specificity

The major structural rearrangement that occurs during substrate binding is the movement of an external loop of polypeptide (residues 99-110) between α-helix D and β-sheet D of the coenzyme-binding domain (Figure 1.5). This loop, known as the active site loop, extends out into the solution in the apo and binary complexes, leaving the active site exposed to the solvent. However, on forming a ternary complex the loop closes over the active site, excluding solvent from the site (Holbrook *et al.*, 1975). The movement of the loop also has the effect of bringing a catalytically important arginine residue (carried on the active site loop) into proximity with the reactants. Arginine-109 serves to stabilise the transition state in the hydride transfer (discussed below).

Loop closure is the rate limiting step in catalysis for L-LDH (Waldman *et al.*, 1987; Clarke *et al.*, 1986) and the conformational change is thought to be mediated through the movement of two 'hinges', connected by a rigid segment, producing one motion (Gernstein and Chothia, 1991).

The second role of the loop is to determine substrate specificity through contacts around residues 101-103. This role has been investigated using site-directed mutagenesis.

While residue 102 is a glutamine in most L-LDHs, there is a third conserved arginine at this locus in the majority of L-MDHs. It was thought that the arginine formed a bifurcated interaction with the carboxymethyl group of oxaloacetate (L-MDH's ketoacid substrate) in the closed ternary complex (Figure 1.6).

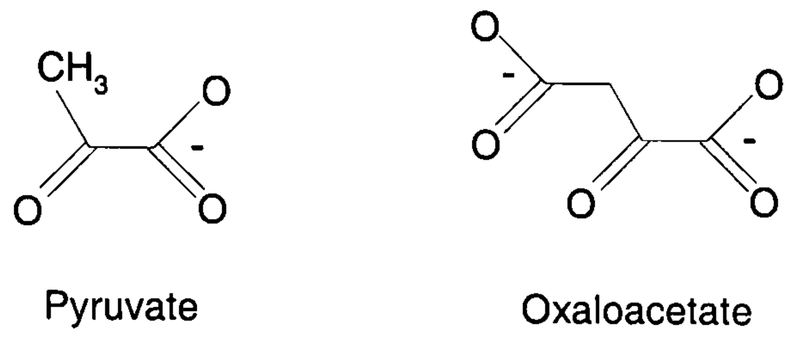


Figure 1.6 The chemical structures of pyruvate and oxaloacetate.

This hypothesis was shown to be correct when the uncharged side-chain of glutamine-102 was substituted for the positively charged side chain of arginine in *Bacillus stearothermophilus* L-LDH (*bs*LDH) (Wilks *et al.*, 1988): The resulting mutant switched its specificity from pyruvate (L-LDH's ketoacid substrate) to oxaloacetate. The catalytic efficiency for pyruvate reduced by 10^4 and increased by 10^3 for oxaloacetate.

Further proof of the active site loop's role in substrate discrimination has come from further site-directed mutagenic studies. For example, rat testicular L-LDH reduces a wide variety of large ketoacids, such as pyruvate ketobutyrate, ketovalerate and phenylpyruvate. The enzyme's broad specificity is thought to arise from its unusually hydrophobic loop residues (methionine-102, valine-103, serine-105) (Clarke and Dafforn, 1998). Inserting this sequence into the loop of *bs* L-LDH caused the mutant enzyme to have a similar ketoacid specificity as the rat enzyme (Wilks *et al.*, 1990).

It should be noted that substrate discrimination is not solely due to the residues on the active site loop and arginine-171. Specificity of an enzyme is determined in conjuction with several other factors, such as charge differentiation, steric clashes, hydrophilic / hydrophobic interactions, etc. and is thus a complex process.

1.4.3 General Reaction Mechanism of NAD^+-dependent L-2-hydroxyacid Dehydrogenases

The chemical reaction mechanisms of the NAD^+-dependent L-2-hydroxyacid dehydrogenases are in principle very similar.

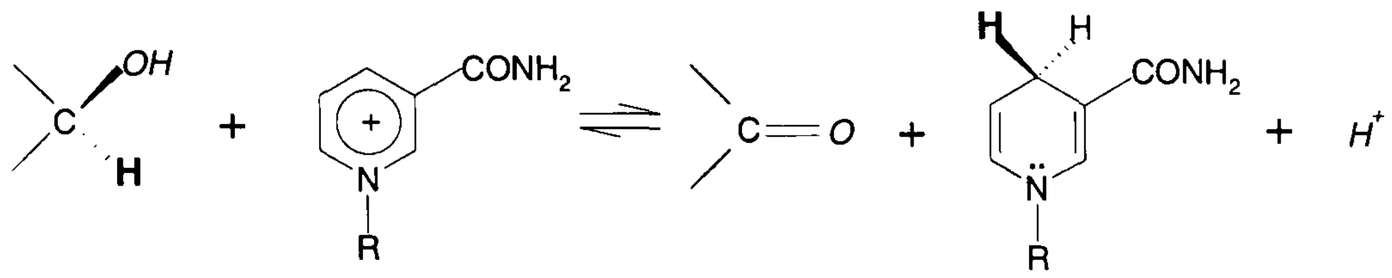


Figure 1.7 Typical mechanism for hydride transfer in $NAD(P)^+$-dependent L-2-hydroxyacid dehydrogenases.

2-Hydroxyacid dehydrogenases have an ordered bi-bi mechanism with the coenzyme binding first followed by the substrate (reviewed in Dalziel, 1975). Typically, at pH 7, the equilbrium constants for the reactions are ~10^4 in favour of the 2-hydroxyacid and NAD^+. However, the binding affinities for both NADH and ketoacid are usually 50-100 -fold stronger than those for NAD^+ and 2-hydroxyacid, respectively.

Hydroxyacid dehydrogenases often exhibit inhibition by excess substrate (Dalziel, 1975). This is due to the formation of an 'abortive' complex of NAD^+ and ketoacid (Figure 1.8). At high concentrations of substrate the ketoacid has a greater chance of forming an adduct with the enzyme-coenzyme binary complex before the NAD^+/H can leave the active site, leading to substrate inhibition.

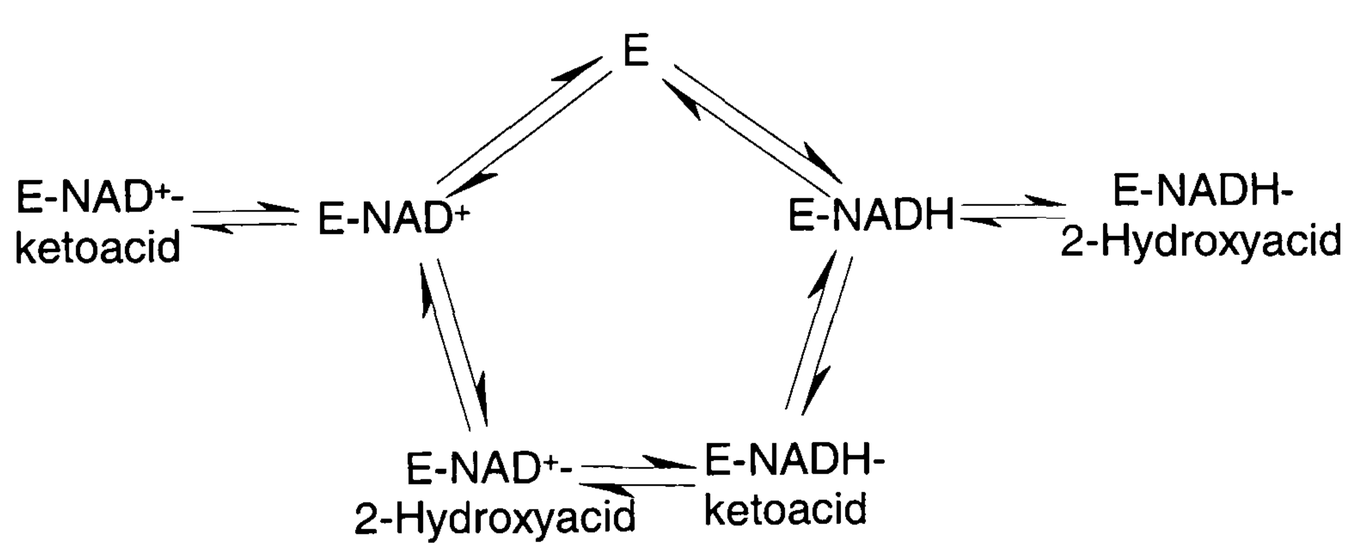


Figure 1.8 The general catalytic cycle of 2-hydroxyacid dehydrogenases. (E= Enzyme).

The Oil-Water-Histidine mechanism proposed by Parker and Holbrook (1977) suggests how the transfer of the hydride ion from the 2-hydroxyacid to NAD^+, may be catalysed. On formation of a ternary complex, the flexible active site loop lowers, closing the active site to bulk solvent (White *et al.*, 1976). This conformational change also pushes the coenzyme further into the enzyme and up against a hydrophobic pocket. This in turn causes the rearrangement of positive charge on the nicotinamide ring from the N1 to the C4 position where the environment is more hydrophilic. The coenzyme's unstable $C4^+$ carbocation is now susceptible to nucleophilic attack by a hydride ion donated by the nearby α-carbon of the 2-hydroxyacid substrate. The hydride exchange is balanced by proton transfer from the C2 hydroxyl group of the substrate to a conserved unprotonated histidine residue (acting as a general acid-base), producing a 2-keto acid. In effect there is a net flux of two electrons from the imidazole ring of histidine to the N1 position of NAD^+.

It was later found that the catalytic histidine in most $NAD(P)^+$-dependent dehydrogenaes is hydrogen bonded to an aspartate (Birktoft *et al.*, 1982). This His-Asp pairing is commonly located in the active sites of many proteins, such as serine proteases (Kraut, 1977). It was suggested that the His-Asp pair acts as a charge-relay system (Blow *et al.*, 1970): The hydrophilically buried aspartate residue's carboxylate group is thought to raise the pK_a of the histidine through the negative charge and also to confine its location (Fersht, 1985). Thus, on binding 2-hydroxyacid substrate to the active site an interface is formed between the hydrophobic environment of the NAD^+ and the hydrophilic environment of

the His-Asp pair. Along with the active-loop closure, this interface is thought to facilitate electron transfer.

The Oil-Water-Histidine model was found to be incomplete; the model did not adequately explain how the unfavourable reverse reaction was catalysed, nor did it consider the role of another conserved arginine (numbered 109 in L-LDH) within the active site: This arginine is distinct from the one which forms a salt bridge with the carboxyl group of the substrate. A revised mechanism was later proposed by Grau *et al.* (1981) (Figure 1.9).

Upon loop closure, the positively charged guanidinium group of the arginine-109, situated on the active site loop, moves to a position between the histidine and 2-oxoacid, thus charge-balancing the active site vacuole. The carbonyl group of the substrate is polarised; electrons being pulled from the α-carbon to the carbonyl oxygen. The histidine residue then donates a proton to the carbonyl oxygen of the 2-keto acid, leaving the α-carbon with a partial positive charge. This carbocation is now susceptible to to hydride attack from NADH. Moreover, the developing charge on the histidine residue would push the arginine away from the substrate, pushing the flexible loop open with the concomitant release of the reaction products.

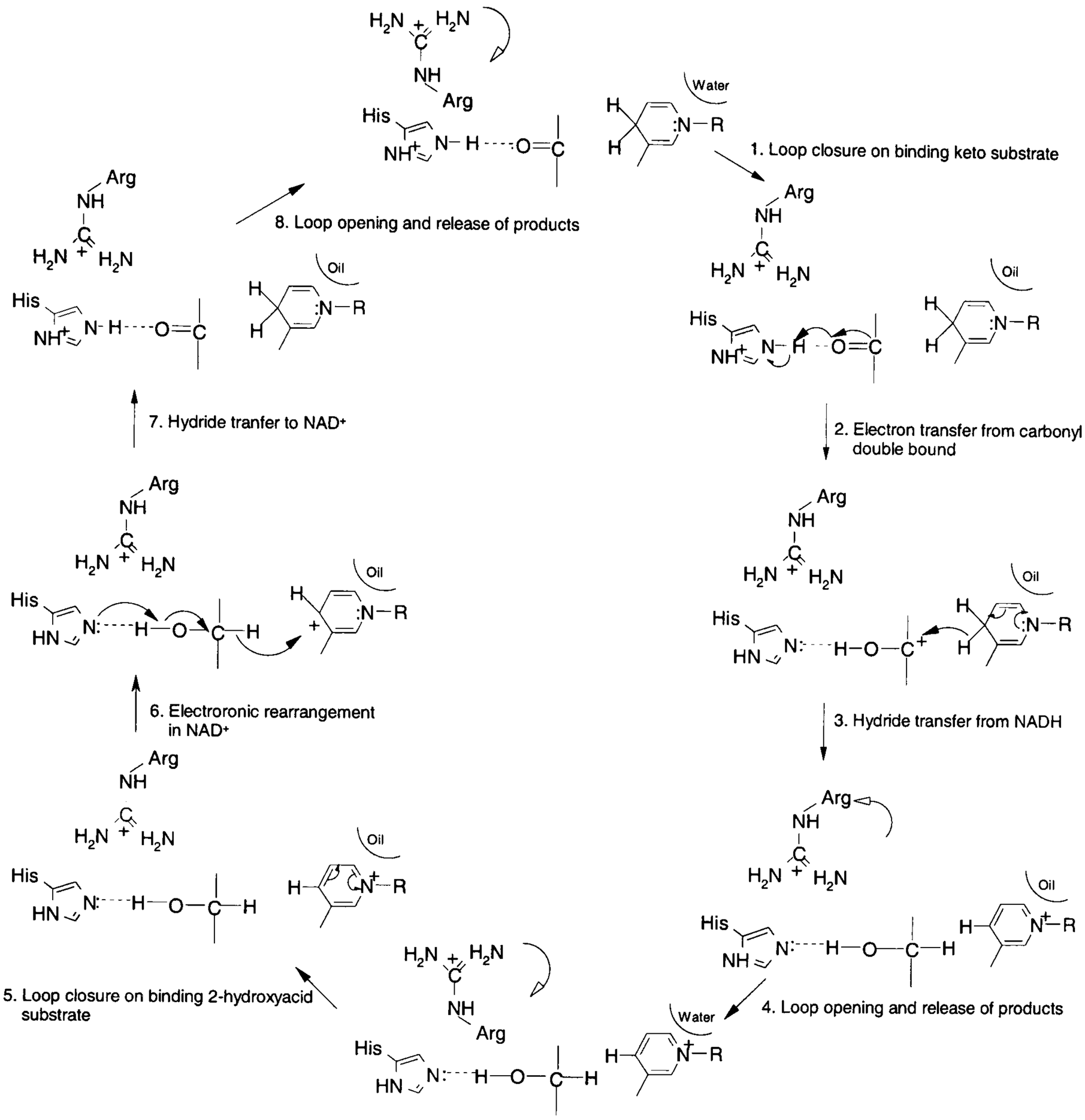


Figure 1.9 The modified Oil-Water-Histidine mechanism proposed by Grau *et al.*, (1981).

The back reaction is essentially the reverse of the mechanism forward reaction: Electronic rearrangements in NAD^+ cause the formation of a positive charge at the C4 position of the nicotinamide ring. NADH is then formed via nucleophilic attack from the C2 hydrogen of the hydroxyl substrate. The histidine regains its positive charge by removing the substrate's O3 hydrogen. This modified Oil-Water- Histidine mechanism thus introduces a balancing of charge repulsion at the nicotinamide ring, and hence, allows for reversibility.

Site-directed mutagenic studies have shown a 400-fold reduction in k_{cat} by the reverse reaction if the catalytic arginine is substituted for the neutral glutamine (Clarke *et al.*, 1986).

1.4.4 Examples of L-2-Hydroxyacid Dehydrogenases

1.4.4.1 L-Lactate Dehydrogenase (discussed in Chapters 6 and 8)

L-Lactate Dehydrogenases catalyse the stereospecific interconversion of pyruate and L-lactate using the NADH/NAD^+ redox cofactor. The enzymes are commonly found in the archaea, prokaryia and eukaryia. In eukaroytes the L-LDH family of enzymes usually exist as tetramers, formed by a tetrahedral arrangement of catalytically independent monomers. All have a very similar overall topology, as originally described for dogfish M_4 L-LDH (Rossmann *et al.*, 1971). In lower organism they are commonly found as dimers (Holbrook *et al.*, 1975).

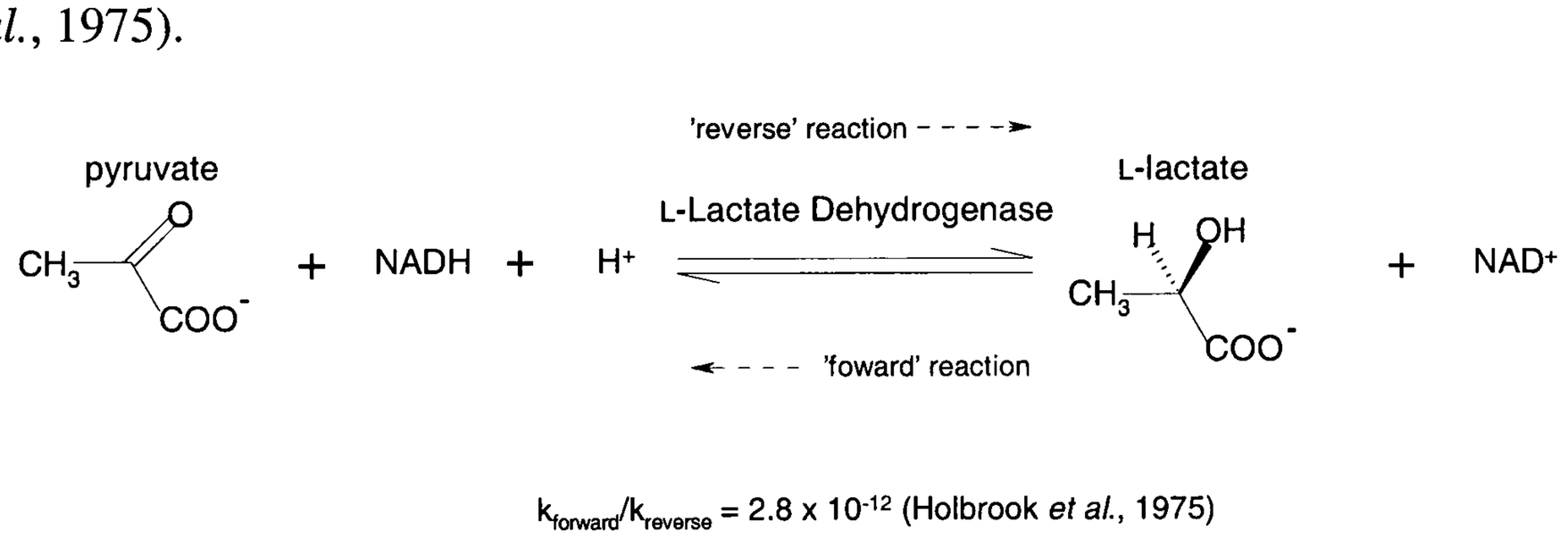


Figure 1.10 The reaction catalysed by L-Lactate dehydrogenase.

The major metabolic role of L-LDH is to allow the continous conversion of glucose to two molecules of lactic acid (which are excreted) and ATP in anaerobic tissues and cells. The enzyme achieves this by reducing pyruvate to lactate, thereby recycling the limited quantities of NAD^+ (NAD^+ is reduced to NADH by glyceraldehyde-3-phosphate dehydrogenase (GAPDH) during the prior metabolic reaction). This anaerobic role is found in micro-organisms and

higher organisms. In the aerobic environments of certain organs of higher animals the reverse reaction is catalysed by L-LDH in order to produce lactate to fuel the heart.

In the reversible reaction the *pro-R* (A-side) hydrogen at C4 of NADH is stereospecifically transferred to the *re* face of pyruvate at C2 to form L- (or *S*-) lactate. The active site residues involved in pyruvate binding and catalysis are shown below (Figure 1.11):-

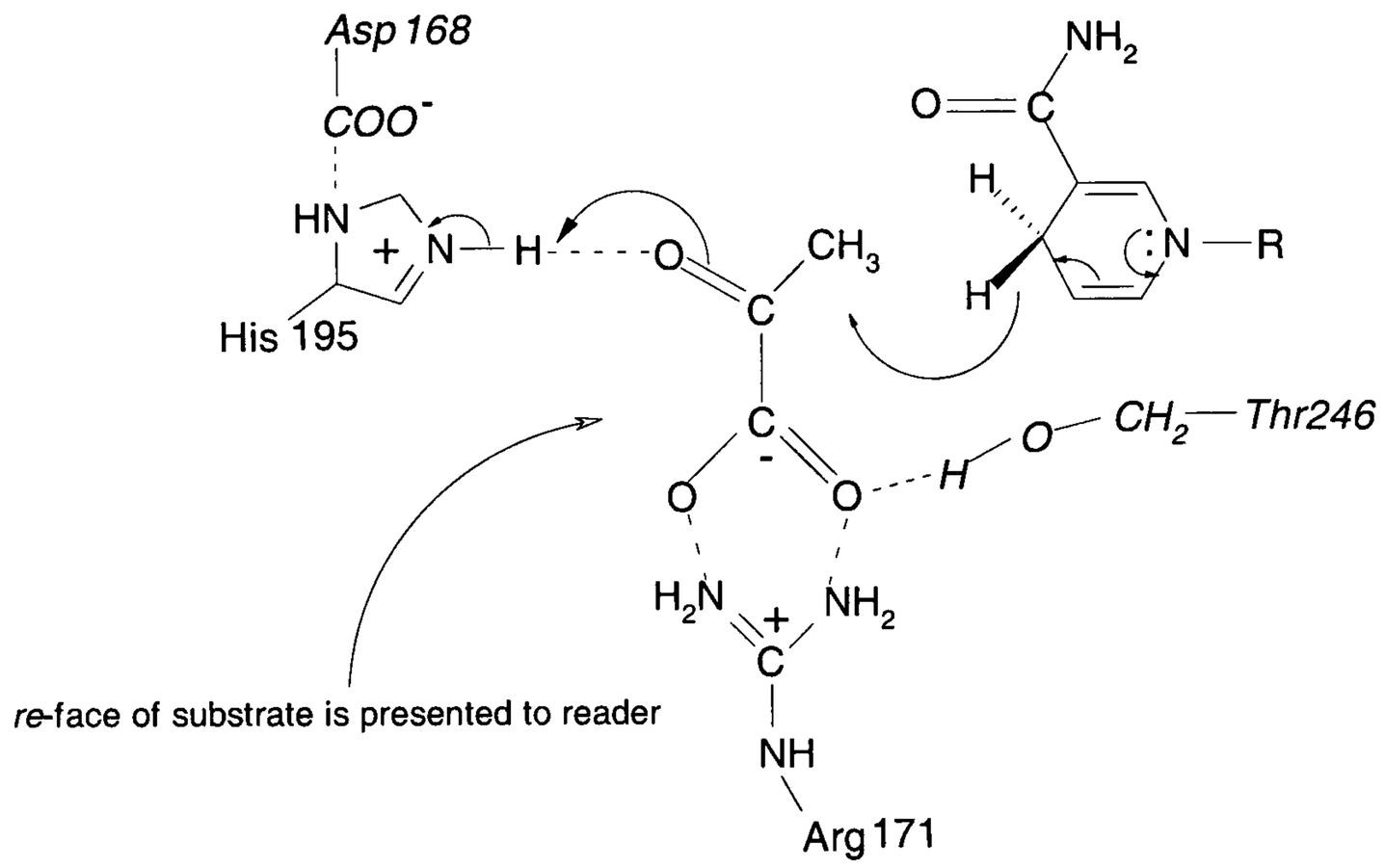


Figure 1.11 The interactions involved in the binding of pyruvate binding and catalysis in the active site of L-LDH.

L-LDH preferably catalyses the reduction of pyruvate, but it can also reduce other small 2-ketoacids with a single negative charge such as 2-oxobutyrate, 2-oxovalerate and 2-oxocaproate (Figure 3.1) to a far less, but still significant degree. In contrast, the enzymatic reaction towards branched substrates like 2-oxoisovalerate and 2-oxoisocaproate is only very marginal (Wilks *et al.*, 1990).

1.4.4.2 L-Malate Dehydrogenase

L-Malate Dehydrogenases (MDH) catalyse the reversible oxidation of L-malate, a four carbon α-hydroxy dicarboxylate, to the ketone oxaloacetate, by transferring a hydrogen from the A-face of NAD^+ (reviewed in Banaszak and Bradshaw, 1975).

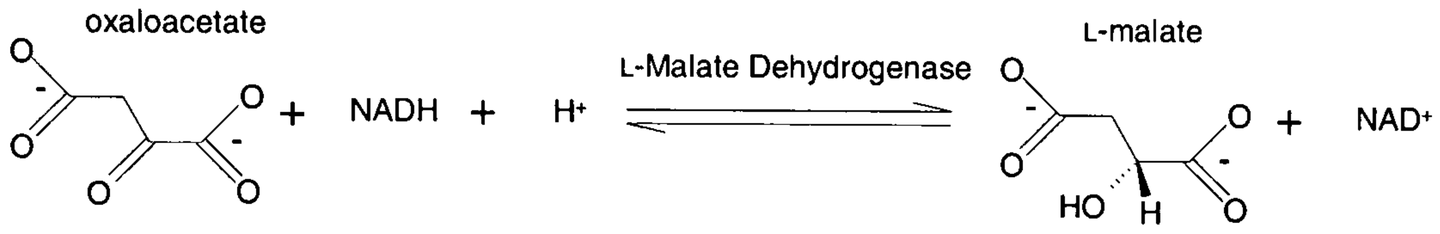


Figure 1.12 The reaction catalysed by L-malate dehydrogenase.

Prokaryotes have only one form of the enzyme and it is located in several cellular compartments. However, in eukaryotes there are two distinct forms (isoenzymes) of dimeric the protein, a cytoplasmic form (cMDH) and a mitochondrial form (mMDH). The physiological role of mMDH is to catalyse the penulimate step of Krebs cycle, whereas the cMDH isoenzyme is involved in the malate-aspartate shuttle (Voet and Voet, 1995).

Even though the two isoenzymes are less than 20% idential in terms of amino acid sequence, they share a common polypeptide chain fold. Furthermore, the structure of MDH has been shown to be remarkably similar to that of L-LDH, with very similar active sites: The mutation of a single active site loop residue in MDH caused a marked shift in specificity and selectivity towards that of L-LDH (Wilks *et al.*, 1988). However, unlike L-LDH, malate dehydrogenase has a very narrow substrate specificity (Boernke *et al.*, 1995). As mentioned earlier, this difference in substrate accommodation is thought to be due the presence of a second conserved arginine residue in MDHs' on the mobile active site loop, also

providing further evidence for a charge-balancing mechanism in 2-hydroxyacid dehydrogenases (Chapman *et al.*, 1999).

1.4.4.3 L-2-Hydroxyisocaproate Dehydrogenase

L-2-hydroxyisocaproate dehydrogenase (L-HicDH) (or broad specificity L-hydroxyacid dehydrogenase) from *L.confusus* was discovered by Schütte *et al.* (1984) during a screen of microbial dehydrogenases. L-HicDH has a broad substrate spectrum, catalysing the reduction of a wide range of L-2-hydroxyacids with branched and unbranched aliphatic side chains of medium size (such as pyruvate, 2-oxoisocaproate and 2-oxocaproate) as well as aromatic residues, such as phenylpyruvate (with *R* containing four to seven carbon atoms) as substrates (Schütte *et al.* 1984; Feil *et al.* 1994) (See Figure 3.1).

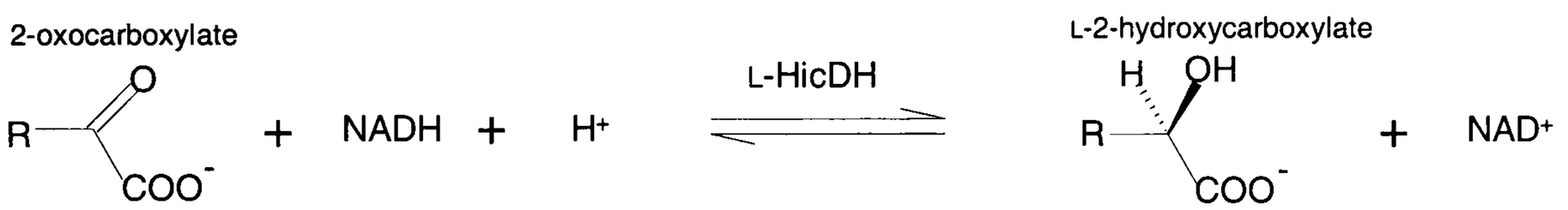


Figure 1.13 The broad specificity reaction catalysed by L-2-hydroxyisocaproate dehydrogenase.

Amino acid alignment studies showed that L-HicDH contains two inserts of two amino acids (residues 102-103 and 106-107) when compared with the L-LDH enzymes (excluding *P.falciparum* L-LDH), in the region corresponding to the active site loop (Feil *et al.* 1994; Niefind *et al.* 1995). As mentioned earlier, this loop is involved in catalysis and in assuring the substrate specificity of the enzymes, and so it was proposed that the inserted residues increase the size and thus the specificity of the substrate-binding region. The location and functional role of these inserted residues were investigated using site-directed mutagenesis (Feil *et al.* 1994). It was demonstrated that even if the four 'inserted' residues were deleted L-HicDH was still active against phenylpyruvate albeit with a

markedly reduced activity. Thus, the four residue insert is not necessary, and therefore, not solely responsible for reactivity towards larger substrates.

It was observed that the deletion of residues Asn106 and Pro107 (Figure 1.4 numbering system) had a marked effect on substrate specificity; the deletions resulted in reduced k_{cat} values for the reactions with pyruvate and 2-oxocaproate. However, a more than 10- fold k_{cat} value increase was observed when 2-phenylpyruvate was used as the substrate. But a more than 100-fold increased k_{cat} value was observed for 2-oxoisocaproate; the catalytic efficiency (k_{cat}/K_M) is 5-fold higher with 2-oxoisocaproate compared to the wild type enzyme. It appears that the deletion of the proline and glutamine residues favours the binding of 2-ketoacid substrate branched at C4.

Further experiments showed that the conserved Gln105 (Figure 1.4) between the two inserts had a comparable substrate-binding role with Gln102 in other L-hydroxyacid dehydrogenases (see 1.4.2.6). Later the crystal structure of *L.confusus* L-HicDH (Niefind *et al.* 1995) revealed that the other conserved glutamine residue between the two amino acid inserts, namely Gln104, also has a crucial function: In most L-LDH structures this residue is directed towards the substrate. This narrows the space that is available for a potential substrate residue, and increases the non-polar character of the region. Steric hindrance from the residue would be a serious problem if attempting to bind branched or aromatic substrates. However, in L-HicDH Gln104 is turned back on itself and fixed by a strong hydrogen bond to Gln115, so that it presents its hydrophobic methylene groups to the substrate binding site and thus cannot disturb the binding of branched or aromatic substrates.

1.5 NAD^+-dependent D-2-Hydroxyacid Dehydrogenases

1.5.1 A Distinct Family of 2-Hydroxyacid Dehydrogenases

D-2-hydroxyacid dehydrogenases catalyse the NADH reduction of 2-ketoacids to give NAD^+ and D-2-hydroxyacid. This family of enzymes includes D-lactate dehydrogenase (D-LDH) (*e.g.* Kochhar *et al.*, 1992), D-2-hydroxyisocaproate dehydrogenase (D-HicDH) (or broad specificity D-hydroxyacid dehydrogenase) (Bernard *et al.*, 1994), glycerate dehydrogenase (GDH) (Greenler *et al.*, 1989), D-3-phosphoglycerate dehydrogenase (D-3PGDH) (Tobey and Grant, 1986) and erythronate-4-phosphate dehydrogenase (Schoenlein *et al.*, 1989).

This enzyme family only emerged recently. Before that the idea that enzymes such as D-LDH was the chiral alternative to the L- enantiomer specific LDH was dismissed. However, Grant (1989) demonstrated that glycerate dehydrogenase and erythronate-4-phosphate dehydrogenase displayed significant sequence homology to D-3-phosphoglycerate dehydrogenase. In contrast, these proteins do not display much similarity with other oxidoreductases or with other 2-hydroxyacid dehydrogenases. The 'creation' of a new family of 2-hydroxyacid dehydrogenases was later reinforced by the discovery of other D- enantiomer specific dehydrogenases.

1.5.2 Amino Acid Sequence Comparison

The amino acid sequence comparison of members of the D-2-hydroxyacid dehydrogenase family are shown in Figure 1.14.

The consensus NAD^+-binding domain $GxGxxGx_{(17)}D$ is found in the middle of the sequence alignment (starting at residue 211 in Figure 1.14). The alignment also shows that the D-HicDH and D-LDH enzymes are related most closely within the family and the erythronate-4-phosphate dehydrogenase is the most distantly related.

The residues implicated in substrate binding catalysis (Arg300) and catalysis (His361; apart from erythronate-4-phosphate dehydrogenase) are highly conserved (Dengler *et al.*, 1997; Kochhar *et al.*, 2000) (see 1.5.3.2). A histidine residue at position 266 (or H205 in the D-HicDH numbering system) is also highly conserved in all members of the D-hydroxyacid family (discussed in Chapter 3).

```
                                                                                1                                                                                
```

1.5.3 Members of the D-Hydroxyacid Dehydrogenase Family

1.5.3.1 *VanH* gene product

With one exception all the D-2-hydroxyacid dehydrogenase family members use NAD^+ and not $NADP^+$ as the coenzyme: The NADPH-dependent D-2-hydroxyacid dehydrogenase is thought to be involved in resistance to the antibiotic vancomycin. In vancomycin susceptible bacterial strains the antibiotic forms a complex with D-alanine - D-alanine dipeptides essential for peptidoglycan crosslinking, thereby disrupting the cell wall of the microorganism. The antibiotic complex is formed through vancomycin hydrogen bonding to the -NH peptide groups of peptidoglycans (Barna and Williams, 1984). However, in the strain *Enterococcus faecium* BM4147, vancomycin resistance is conferred by five gene products (Wu *et al.*, 1995). One of these genes, *vanH*, encodes a NADPH-dependent D-2-hydroxyacid dehydrogenase. It has been suggested that the molecular basis of vanomycin resistance is the formation of D-lactate - D-alanine depsipeptides which the antibiotic is incapable of recognising (Arthur *et al.*, 1992).

1.5.3.2 D-Lactate Dehydrogenase

Most of the biochemical and kinetic information regarding the D-2-hydroxyacid dehydrogenase family has come from the study of D-LDH.

Unlike the ubiquitous L-LDH enzyme, D-lactate dehydrogenase (D-LDH) is only expressed in lower organisms (anthropods, mulluscs, etc.) and bacteria: Some micro-organisms, such as *Lactobacillus plantarum*, possess both the D- and L-LDH enzymes (Vinals *et al.*, 1993.). Like most D-hydroxyacid

dehydrogenases, D-LDH is a dimeric protein with each subunit consisting of a substrate-binding domain and a cofactor-binding Rossmann fold.

A number of crystallographic, site-directed mutagenic, chemical modification, kinetic and physiochemical property studies have been carried out on D-LDH's from various sources (Dym *et al.*, 2000; Vinals *et al.*, 1995; Kochhar *et al.*, 1992; Kochhar *et al.*, 2000). These investigations show that the active sites of the D- and L-LDH enzymes are remarkably similar: In both enzymes, a histidine-aspartate (or glutamate) couple appears to act as a H^+ donor / acceptor and a conserved arginine residue binds the α-carboxylate group of the substrate.

As mentioned earlier, primary and tertiary structural analysis of the D-LDH enzymes have also revealed the presence of a conserved aspartate some 17 residues up stream of the glycine -rich (G/AXG/AXXG/A) putative NAD^+-binding motif.. As with all NAD^+-dependent L-2-hydroxyacid dehydrogenases, site-directed mutagenic studies on *Lactobacillus delbruekii subsp. bulgaricus* D-LDH have shown that the conserved aspartate residue is essential for $NAD^+/NADP^+$ discrimination: Substituting D175 for an alanine causes a 40-fold shift in coenzyme preference from NADH to NADPH (Bernard *et al.*, 1995).

The presence of analogous amino acids involved in substrate binding and catalysis (in virtually identical positions in the two groups of proteins) but lacking any significant sequence identity seems to suggest an apparent evolutionary convergence of the active sites (and cofactor-binding sites) of L-LDH and D-LDH, and all D-2-hydroxyacid deydrogenases. This in turn seems to suggest that the two families of enzymes share a similar reaction mechanism.

The structural basis for the opposing enantiospecificities of D-LDH and L-LDH has been the subject of a number of biochemical studies (*e.g.* Siebanaller *et al.*, 1983). Following the crystallographic elucidation of the structures of both enzymes (Dym *et al.*, 2000; Dunn and Holbrook, 1996; Grau *et al.*, 1981), the ability of LDH to bind only one stereoisomer of lactate could be easily explained in terms of steric restrictions at the active sites of the enzymes. The same catalytic mechanism can be used to reduce a ketoacid to an L- or D-hydroxyacid if the spatial relation of the catalytic histidine ring and nicotinamide ring are kept constant and the 2-ketoacid is rotated through 180° so that the hydride approaches from *re*- or the *si*-face of the ketone . The reduction to either L- or D-hydroxyacid is achieved by arginine which binds the α-carboxyl group of the substrate being rotated by 180° (*cf.* Figures 1.11 and 1.15).

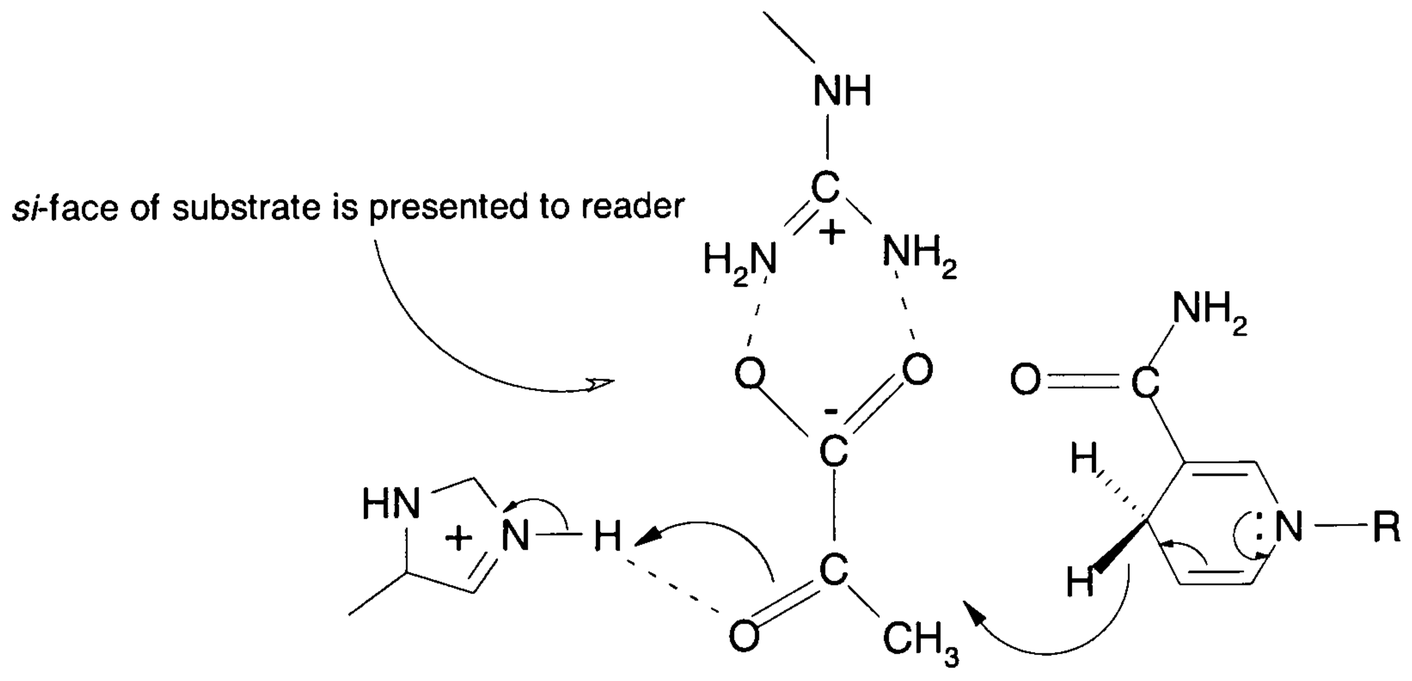


Figure 1.15 The interactions involved in the binding of pyruvate binding and catalysis in the active site of D-LDH.

In L-LDH (and all L-2-hydroxyacid dehydrogenases) transfer of the C-4 hydrogen of the dihydronicotinamide ring is from the *re* face of the substrate, forming the L-isomer of the hydroxyacid. Whilst in D-LDH (and presumably in all D-2-hydroxyacid dehydrogenases) the hydrogen is added to the *si* face to

form the D-isomer. In both LDH enzymes the A-side hydrogen of NAD is transferred.

The basic understanding of the molecular principles behind the enantioselectivity in D- and L-2-hydroxyacid dehydrogenases led to attempts at engineering D-LDH activity into L-LDH. The strategy employed by Sakowicz *et al.* (1995) was to induce the pyruvate molecule in *Bacillus stearothermophilus* L-lactate dehydrogenase (*bs*LDH) to bind in an alternative, 180°-rotated, orientation that would lead to D-lactate production. In order to achieve this Arg-171, the carboxylate binding residue of the L-enzymes, was mutated to a tyrosine and Gln-102, on the active site loop, was mutated to an arginine in order to form a new Arg-COO^--binding interaction (Figure 1.16).

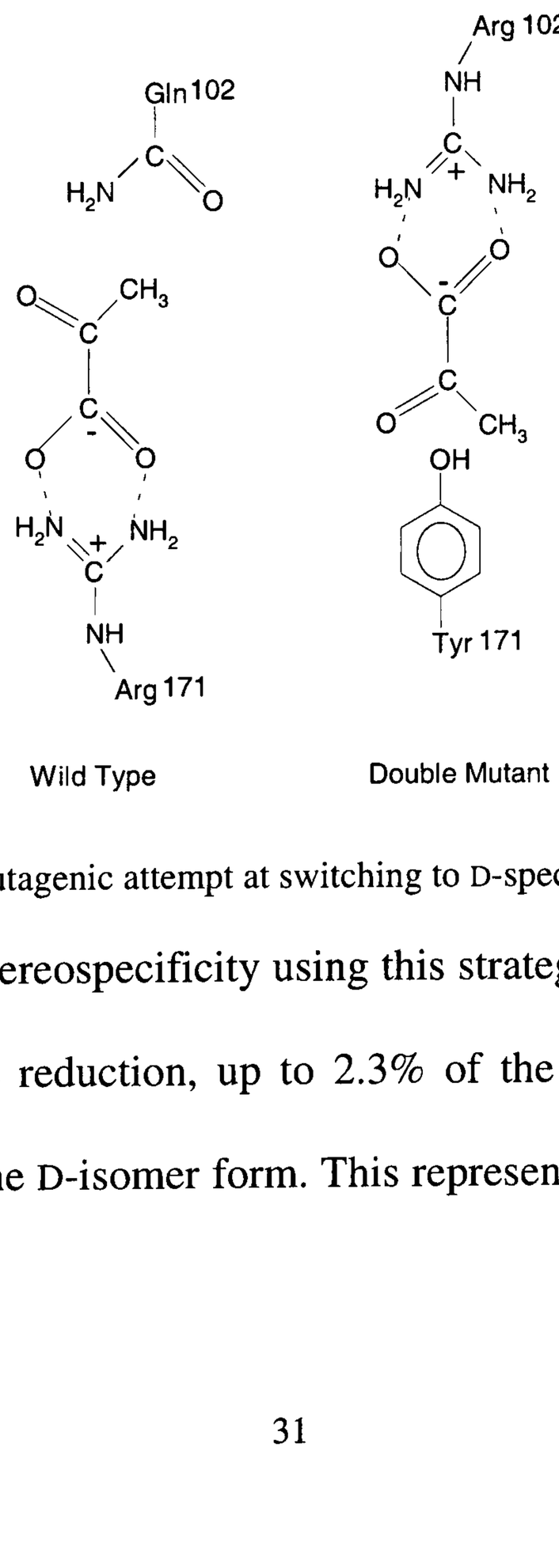


Figure 1.16 Site-directed mutagenic attempt at switching to D-specificity in *bs*LDH.

Switching substrate stereospecificity using this strategy met with a measure of success; upon pyruvate reduction, up to 2.3% of the lactate produced by the double mutant was of the D-isomer form. This represents an estimated ~10^5-fold

relaxation of the natural L-stereoselectivity of *bs*LDH (Anderson and La Reau, 1988).

The precise physiological role of D-LDH is as yet uncertain, however, like one L-LDH, D-LDH might simply catalyse the final reaction in anaerobic glycolysis. This is strongly supported by the fact that arachnids possess only the D-isomer specific LDH (Long and Kaplan, 1968).

Another possible metabolic function for D-LDH is the detoxification of methylglyoxal (Figure 1.17). Methylglyoxal is formed as a by-product in NADPH-dependent acetone and acetol monooxygenase reactions, and it is known to be a potent inhibitor of mitosis (Casazza et al., 1984).

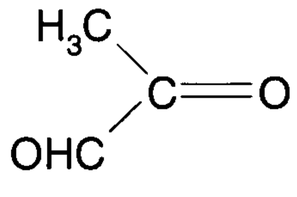


Figure 1.17 The chemical structure of methylglyoxal.

It has been known for sometime that glyoxalase I and II tandemly converts methylglyoxal to D-lactate (Douglas and Shinkai, 1985). So, it has been suggested that D-LDH converts the D-lactate into the metabolically versatile pyruvate substrate.

1.5.3.3 D-Glycerate Dehydrogenase (or Hydroxypyruvate Reductase)

D-glycerate dehydrogenase catalyses (GDH) the NADH-linked reduction of hydroxypyruvate to D-glycerate (Figure 1.18) in a diverse range of organisms; the enzyme has been purified from bacterial, plant and mammalian sources (Goldberg *et al.*, 1994; Greenler *et al.*, 1989; Rosenblum *et al.*, 1971). In all cases the GDH

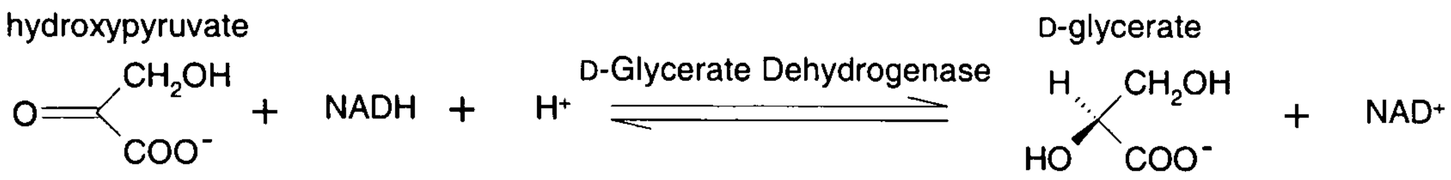


Figure 1.18 The reaction catalysed by D-glycerate dehydrogenase.

enzymes are dimeric in nature with a monomer molecular mass between 33,000 and 42,000 Daltons.The GDHs' belong to the A class of NADH-linked dehydrogenases with the hydrogen being transferred to the *si* face of hydroxypyruvate (Goldberg *et al.*, 1994).

GDH is one of the enzymes involved in the serine cycle. This cycle is one of the ways in which type II methyltrophs incorporate inorganic carbon (Brock and Madigan, 1988).

1.5.3.4 D-3-Phosphoglycerate Dehydrogenase

D-3-phosphoglycerate dehydrogenase (D-3PGDH) is involved in the 'committing step' in serine biosynthesis (Tobey and Grant, 1986). Serine biosynthesis is important in that it not only controls the concentrations of serine in an organism, serine is also a intermediate in the synthesis of cysteine and glycine (Figure 1.19).

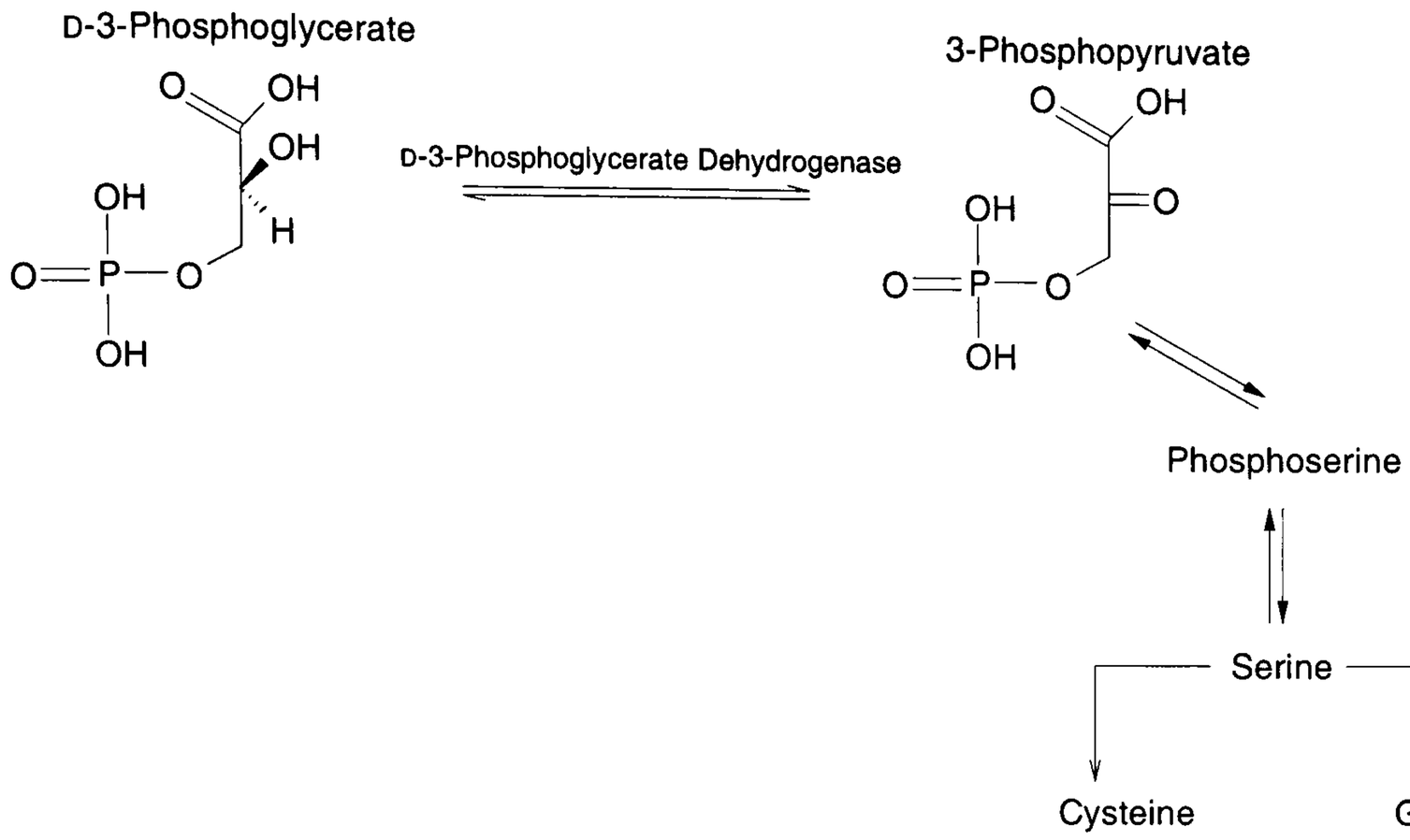


Figure 1.19 Catalytic steps in the pathway of serine biosynthesis.

D-3PGDH is unusual among the class of D-2-hyroxyacid dehydrogenases in that it shows a marked allosteric regulatory response: Serine binds cooperatively to the enzyme and inhibits its catalytic activity. The inhibition is brought about by a reduction in the maximal catalytic rate of the enzyme rather than through a weakening of substrate interaction (Clarke and Daffron, 1998). Also, unlike the other members of the enzyme family, D-3PGDH has three domains, and it is a tetramer, rather than the conventional D-family dimer. It is thought that the extra domain has a serine binding site which mediates the regulatory response (Schuller *et al.*, 1995).

1.5.3.5 Erythronate-4-Phosphate Dehydrogenase

The DNA sequence of a gene essential for pyridoxine (Vitamin B_6) biosynthesis, *pdxB*, appears to be rather similar in sequence to D-3-phosphoglycerate dehydrogenase (Schoenlein *et al.*, 1989). Schoenlein *et al.*, (1989) have also suggested that 4-hydroxy threonine is a major intermediate in pyridoxine biosynthesis. Therefore, it is thought that the *pdxB* gene product encodes for a dehydrogenase that reduces the hydroxyacid erythronate-4-

phosphate to form 3-hydroxy-4-phospho-hydroxy 2-ketobutyrate, an intermediate thought to be required for 4-hydroxy threonine biosynthesis (Figure 1.20).

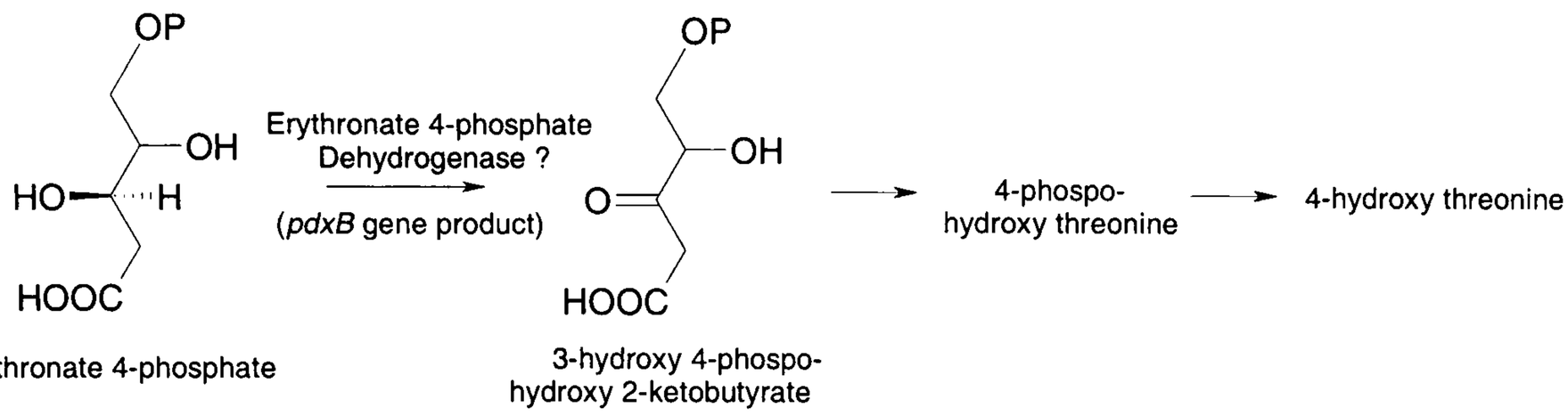


Figure 1.20 Putative steps in the production of 4-hydroxy threonine.

1.5.3.6 D-2-Hydroxyisocaproate Dehydrogenase (discussed in Chapter 3)

D-2-hydroxyisocaproate dehydrogenase (D-HicDH) (or broad specificity D-hydroxyacid dehydrogenase) has a broad substrate spectrum (similar to that of L-HicDH) (Bernard *et al.*, 1994). D-HicDH catalyses the reduction of a wide range of D-2-hydroxyacids with branched and unbranched aliphatic side chains of medium size. The physiological functions of the enzyme are as yet unknown.

Jensen (1976) has suggested that macromolecular evolution is a result of substrate ambiguity in enzymes: The primitive enzymes were likely to have broad substrate specificities. It is possible to imagine the evolution of a protein from a broadly specific protein (such as D-HicDH or L-HicDH) to a narrowly specific protein that is committed to a particular pathway. Events such as gene duplication would allow the proteins to diverge from their ancestral molecule.

Chapter 2

Crystallographic Techniques

2.1 Protein Crystallisation

2.1.1 Introduction

One of the main obstacles in obtaining a crystallographic macromolecular structure is the ability to grow single crystals that are suitable for high resolution X-ray diffraction analysis.

Growing good quality macromolecular crystals is still essentially a trial-and-error process where the factors that effect crystal growth are systematically sought: These variables are then optimised to produce the best crystals possible.

Crystallisation screens are commercially available (e.g. Hampton Research, Laguna Niguel, CA92677-3913, USA) and were employed to search for initial crystallisation conditions. The Hampton Research protein screens are a sparse matrix (Jancarik and Kim, 1991) of 98 conditions (50 solutions in screen 1 and 48 in screen 2) that have proven to be successful for crystallisation in the past.

2.1.2 Principles of Protein Crystallisation

Protein crystallisation can be described as a process whereby the protein is slowly precipitated from its solution. It is a complex process that is as yet poorly understood for several reasons: Firstly, proteins are structurally dynamic molecules whose physiochemical properties (such as solubility) vary as a function of innumerable factors such as pH, ionic strength, solvent content, contaminants, temperature, etc. Added to this is the inadequate understanding of the forces which drive macromolecular crystallisation from a solution.

For crystallisation to occur the protein solution, typically greater than 95% pure, must be brought to supersaturation. Protein molecules aggregate to form small nuclei at a high level of supersaturation (any higher, and the protein would aggregate as an amorphous precipitate). Nuclei formation should then occur slowly to maximise the degree of order in the structure.

Ideally, the protein solution, by depletion, would return to the lower end of supersaturation where the nuclei would continue to grow into defect free crystals.

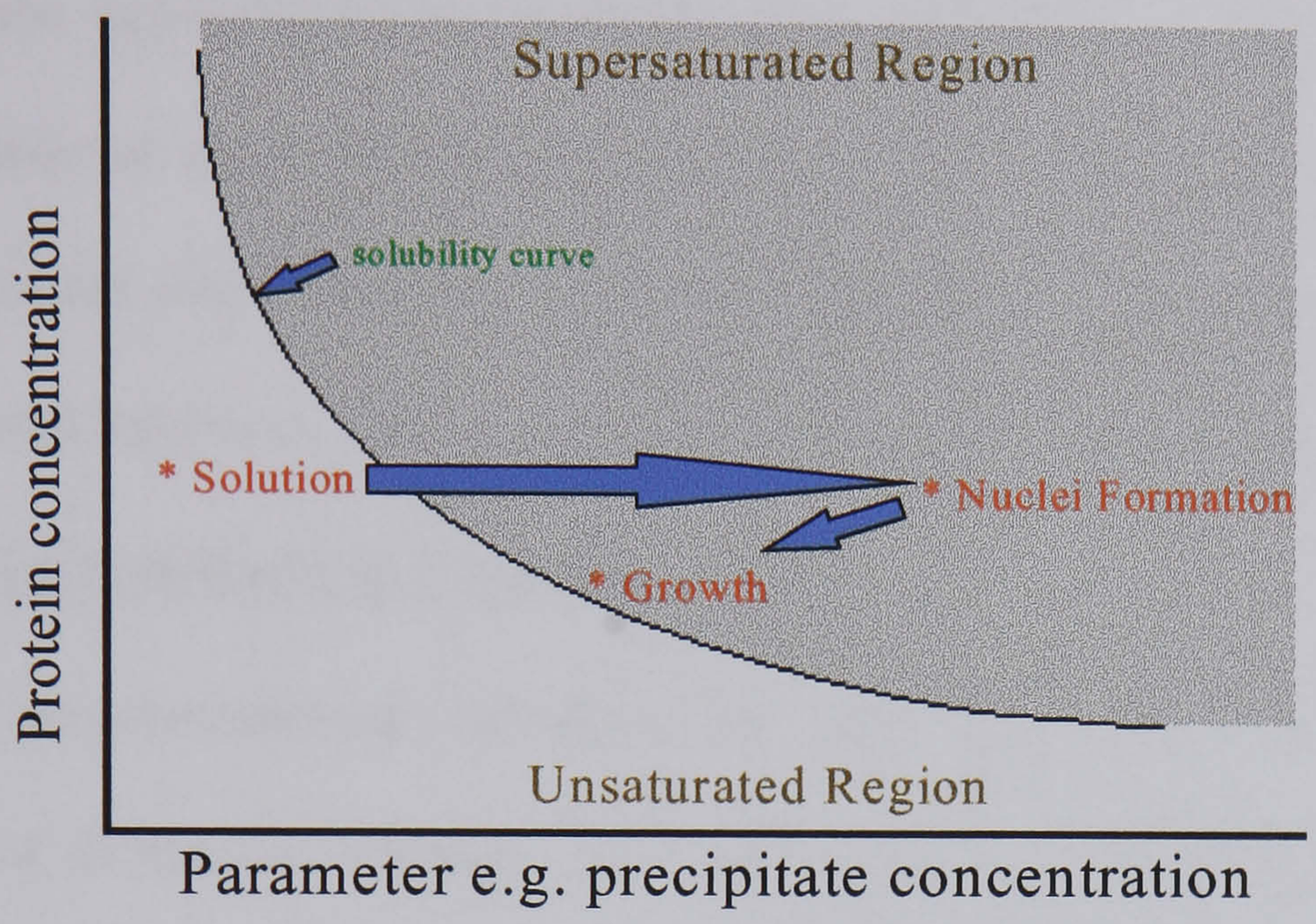


Figure 2.1 Solubility curve for a typical protein.
Maintaining a high level of supersaturation would result in the formation of many nuclei and thus many small crystals.

2.1.3 Crystallisation Techniques

There are four main crystallisation techniques that can be used; dialysis, liquid-liquid diffusion, batch crystallisation and vapour diffusion.

All of the crystallisation trials carried out used either the hanging or sitting drop variant of the vapour diffusion method (McPherson, 1990). In the hanging drop method up to 5µl of protein solution is placed on a plastic coverslip and is usually mixed with the same volume of precipitant solution to make up the mother liquor. The coverslip is then placed upside down over a well containing about 1ml of precipitant solution. The well is sealed by applying oil or vacuum grease around the edge of the depression before the coverslip is put into place. Disposable plastic tissue culture plates were used to supply the wells. The sitting drop method is similar, however, up to 100µl of the protein solution mixture is

pipetted into the depression of a micro-bridge (Hampton Research) that has been placed in the well itself.

In the vapour diffusion methods of crystallisation, equilibration proceeds by diffusion of water between the two solutions. This leads to the concentration of the solutes in the drop due to the vast differences in the volumes of the protein and well solutions. This in turn leads to supersaturation of the protein solution.

2.1.4 Cryo-cooling Crystals

All macromolecular crystals suffer from dose-dependent radiation induced loss of diffraction intensity. The X-ray induced damage is due to presence of large amounts of water in the protein crystal and the potential diffusion of free-radicals within it (Henderson, 1990). This problem can be dramatically reduced by flash-cooling the crystal in the presence of a cyro-protectant to very low temperature (Garman and Schneider, 1997).

Lowering the temperature of the crystal gradually would result in the formation of ice. Ice effects the protein diffraction pattern by contributing to the diffraction pattern and by shearing the protein crystal lattice. Cryoprotectants are chemicals, such as alcohols or sugars, which reduce the formation of ice. Flash-cooling a crystal (taking it from non-cryo conditions to cryo conditions rapidly) with a cryo-protectant causes the aqueous solution, in and around the crystal, to freeze amorphously, like glass. This process, known as vitrification, can sometimes lead to an increase in mosaicity (the degree of order in the crystal lattice). However, there are several advantages of cryo-cooling crystals such as the potential of collecting a data set from a single crystal due to reduced radiation damage; whereas, data collection at room temperature would normally

require multiple crystals, which may lead to problems when scaling the multiple data sets together. Also, a higher resolution data set might be obtained from a crystal that had been flash-cooled as compared to one mounted at room temperature, since protein damage due to radiation is reduced and thermal vibrations within the crystals are reduced at low temperature.

The methods employed in the mounting and cooling of crystals prior to cryogenic data collection involved soaking the protein crystals in mother liquor containing increasing concentrations of the cyro-protectant additive (often with a slight increased precipitant concentration). The crystals were then suspended in a small nylon loop (surrounded by a thin layer of cryo-protectant solution) and rapidly plunged into a stream of gaseous nitrogen set at a constant 100K: The temperature of which was controlled by a Cryostream (Oxford Cryosystems).

2.2 X-ray Sources and Detectors

2.2.1 Introduction

Crystallographic protein structure determination requires an X-ray source and an X-ray detector to determine the intensity and position of the diffracted beam. During this work either an in-house rotating anode generator with a MacScience double image plate or synchrotron radiation with a Mar Research single image plate was used for data collection purposes.

2.2.2 X-ray Sources

X-rays are electromagnetic radiation with wavelengths of 1000 - 0.1Å. There are two main types of X-ray source commonly used; the home laboratory rotating anode generator and the more powerful synchrotron radiation source.

2.2.2.1 Rotating anode generators

In the rotating anode generator electrons from a heated tugsten filament are accelerated through a potential difference towards a rotating cylindrical metal anode under a vacuum. The anode target can be made from molybdenum, cobalt, iron or some other metal: For protein crystallography the anode is usually copper.

Most of the electron energy is lost as heat when the particles strike the anode. However, some of the energy is converted into X-ray radiation, this can be emitted in two distinct ways. One way is as a smooth continuous intensity distribution as a function of wavelength (Figure 2.2). This is because charged particles (such as electrons) emit electromagnetic radiation, known as 'Bremsstrahlung', when they are either accelerated or decelerated. The cut-off at the shorter wavelength, indicating the most energetic X-rays, is where a photon will have energy corresponding to all the energy of one electron.

If the potential difference is great enough then a point will be reached when the impacting electrons will cause electronic transitions between the inner orbitals in the atoms of the anode material: Electrons in the anode atoms shoot out of lower lying orbitals, this is then accompanied by electrons from higher orbitals occupying the empty orbitals and releasing energy in the form of X-rays at a specific wavelength. This gives rise to what is known as characteristic radiation, as the wavelength of the emitted X-ray is dependent on the anode metal type. It is manifest as sharp intense peaks in the spectrum.

If copper is used as the anode characteristic X-ray radiation will be emitted at 1.54178Å and 1.39217Å. The former X-ray wavelength, called CuK_{α} radiation,

comes from a L to K-shell electronic transition. This is the most intense radiation peak emitted and was the X-ray wavelength used during in-house data collection.

X-rays emitted at 1.39217Å, called $CuK_{ß}$ radiation, are due to the larger M to K-shell transitions.

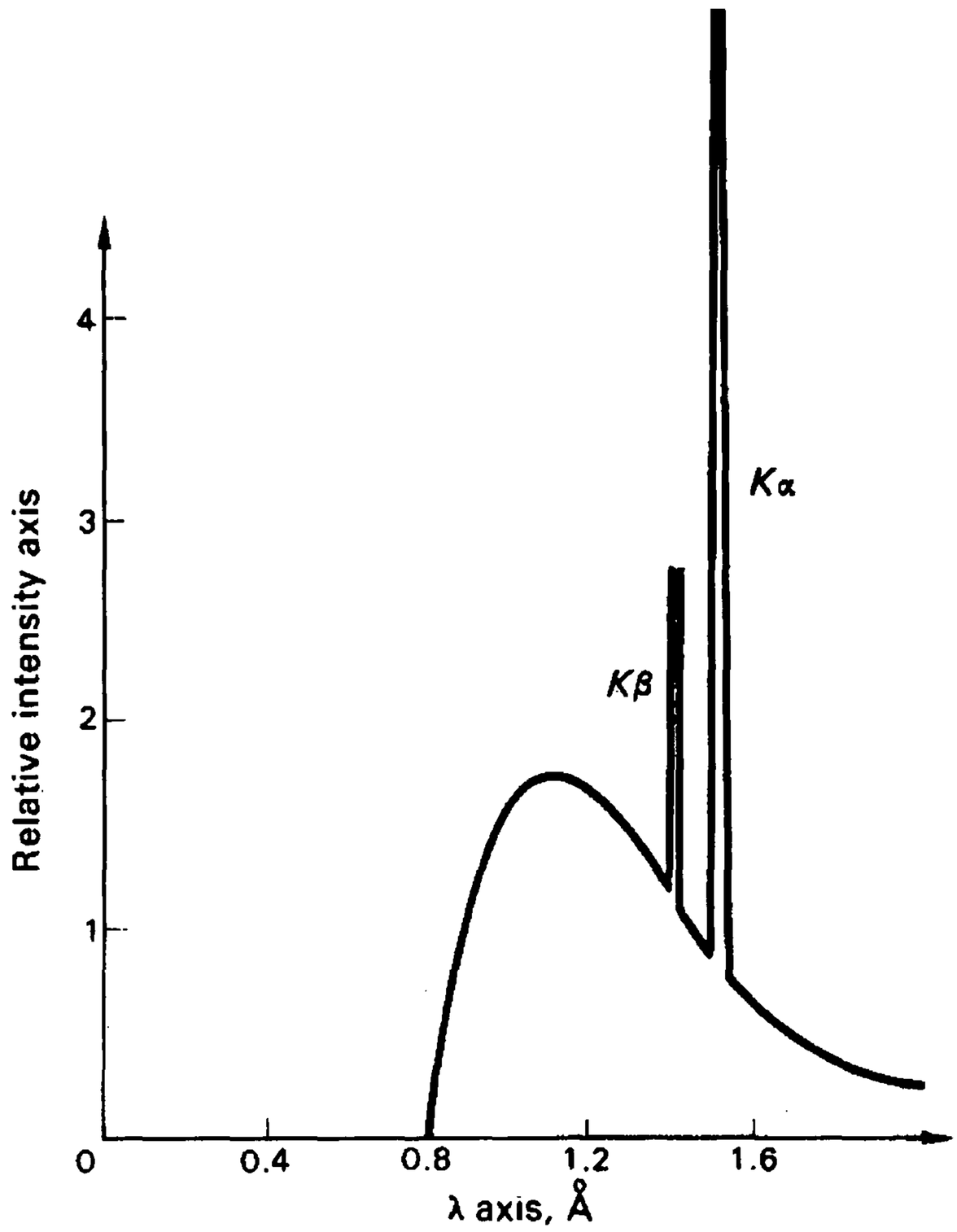


Figure 2.2 X-ray spectrum from a rotating anode / sealed tube with copper as the anode. The X-ray intensity of the emitted radiation is on an arbitrary scale. (From Blundell and Johnson, 1976).

Earlier generators had a fixed piece of metal as the anode (sealed tubes), but, due to heating of the anode caused by the electron beam at the focal spot there had to be a limit on the maximum power of the sealed tube. However, as the electron beam is not continually focused on the same spot in the rotating anode tube the heat can be more efficiently removed by water cooling: This ensures

that the rotating anode can run at a higher power than the sealed tube and thus produce more intense X-rays.

In most diffraction experiments monochromatic X-rays are required. If a rotating anode tube is used with a copper anode the high intensity CuK_{α} radiation can be selected for by using either a nickel filter or a piece of graphite. The nickel filter would remove most of the $CuK_{ß}$ radiation and some of the continuous spectrum. But the intensity of the K_{α} radiation would also be slightly reduced. Graphite, however, reflects CuK_{α} X-rays at a scanning angle of 13.1° (Drenth, 1995) and produces 'cleaner' X-rays with a narrower wavelength bandwidth.

2.2.2.2 Synchrotron radiation

It is preferential to use a synchrotron radiation source as they give many orders of magnitude more intense X-rays than conventional in-house generators. The first synchrotrons were built to accelerate electrons and positrons to high energies (nearly to the speed of light) for experiments in particle physics.

'Bending magnets' are used to ensure that the particles travel in an approximately circular motion. As the charged particles pass through the bending magnets they are accelerated towards the centre of the ring. Therefore the particles lose energy by emitting electromagnetic radiation. During the 1970s crystallographers started to utilise the intense X-ray radiation 'by-product' and by the 1980s radiation-dedicated synchrotrons had been built.

In addition to bending magnets the radiation-dedicated synchrotrons also have wigglers, wavelength shifters and undulators in the straight sections (Helliwell, 1992). These are magnetic devices which cause the charged particles to

accelerate (or decelerate) causing X-ray emission but give no net displacement in particle trajectory.

Monochromators for synchrotron radiation are usually made from a pair of germanium or silicon crystals and have a narrower wavelength bandwidth than graphite.

There are several advantages of using synchrotron radiation over a home laboratory source. For example, the high intensity X-rays usually enable diffraction data to be collected to a higher resolution. Also, a brighter source ensures that better data can be collected from weakly diffracting crystals (such as small crystals or crystals with very large unit cells). The X-ray beam at a synchrotron has a low divergence resulting in sharper reflections (diffracted spots) and thus a better signal-to-noise ratio.

Another advantage of synchrotron radiation is its wavelength tunability. Any wavelength within the spectral range can be selected for by adjusting the scanning angle of the monochromator; this property is of particular use in the multiwavelength anomalous dispersion phasing method.

2.2.3 X-ray Detectors

2.2.3.1 Introduction

Photographic film used to be relied upon to record diffraction data. Collecting and analysing such data was a time consuming process as hundreds of rolls of film had to be developed and then the density of the spots individually measured with a densitometer. However, in recent years faster, more accurate X-ray detectors that allow data collection to be automated have been developed: These

include electronic & charge coupled device (CCD) area detectors (Pflugrath, 1992) and image plates.

2.2.3.2 Image plates

Image plates are composed of a flat base coated with a 150μm thick layer of BaFBr phosphor (a very efficient absorber of X-rays) doped with europium (Eu^{2+}) (Amemiya and Miyhara, 1988). Irradiation of the image plate by X-rays cause the Eu^{2+} ions to be excited into quasi-stable Eu^{3+} ions. This stored energy can be released by scanning the image plate with a He-Ne red laser; the laser stimulates the Eu^{3+} ions into emitting violet light. The intensity of the emitted light is proportional to the number of X-ray photons absorbed by that region of the plate. Thus, by filtering away the incident red light and measuring the violet light with a photomultiplier, it is possible to determine the X-ray intensities of the spots on the plate. Any residual 'image' can then be erased by exposing the image plate to bright yellow light.

Fading of the stored image is not a problem in X-ray diffraction experiments as it takes about two months for a 50% reduction in image intensity (Amemiya and Miyhara, 1988).

Apart from being faster and easier to use, image plates have a dynamic range ($1:10^5$) several orders of magnitude larger than photographic film.

A double image plate (DIP) was employed for in-house data collection. This allows data to be collected rapidly; whilst one image plate is irradiated by X-rays the other is being scanned and then erased ready for the next round of irradiation.

2.3 X-ray Diffraction Theory

2.3.1 The Crystal

The basic repeating unit of a crystal is known as the unit cell: When the unit cell is repeated through translations parallel to the three edges (x, y, z) defined by the crystal axes, without rotation, it gives the crystal lattice. Therefore, a crystal can be regarded as a semi-infinite three-dimensional array of unit cells.

The unit cell is defined by three vectors **a, b** and **c** (in the x, y and z directions, respectively) with the angles α, ß and γ between them.

In most crystals, the parallelepiped-shaped unit cell has internal symmetry in that two or more molecules are related by axes or planes of symmetry. The largest aggregate of molecules that when repeated an integral number of times through symmetry operations yield the entire unit cell is known as the asymmetric unit (Figure 2.3). For protein molecules, there are only 65 possible types of crystal lattice (ie. space groups) in which asymmetric units may crystallise.

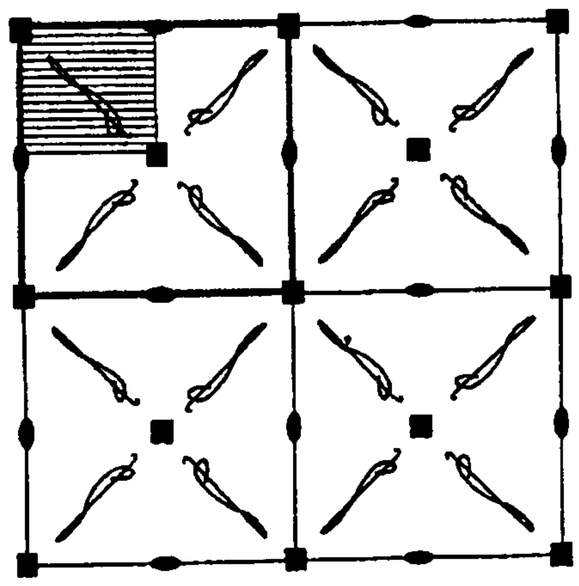


Figure 2.3 An example of a unit cell and its asymmetric unit. A single unit cell is outlined heavily and the asymmetric unit is shaded.

2.3.2 X-ray Diffraction

Diffracted beams are commonly referred to as 'reflections'. This is because in the Bragg model of X-ray diffraction each spot in the diffraction pattern is seen as a reflection from a set of imaginary equidistant parallel planes which pass through atoms within the crystal.

The planes cut an axis, for example the *x*-axis, into equal parts that have a length of a/1, a/2, a/3,etc. The integers 1, 2, 3 etc. are known as indices. If a set of planes cut the *x*-axis into a/h, *y* into b/k and *z* into c/l pieces, they are known as the (*h k l*) planes. If a set of planes are parallel to an axis that particular index is 0.

A set of planes (*h k l*) only produce a diffracted beam when the incident X-ray waves, of wavelength λ, make an angle θ with the planes, such that θ meets the conditions of Bragg's law:

$$2d\sin\theta = n\lambda$$

where *d* is the interplanar spacing and n is an integer.

Consider diffraction from a two-dimensional array of scatterers (Figure 2.4). A number of equidistant parallel lines can be drawn through all scatterers. According to the Bragg model of diffraction, each 'plane' acts as a reflecting surface for X-rays. If the interplanar distance is an integral number of wavelengths, ie. $2d\sin\theta = n\lambda$, then all emerging X-rays will be in phase and thus interfere constructively to produce a strong diffraction beam. However, incident X-rays at any other angle would cause the reflected beams from successive planes to be out of phase; these would destructively interfere with one another and produce no diffracted beam.

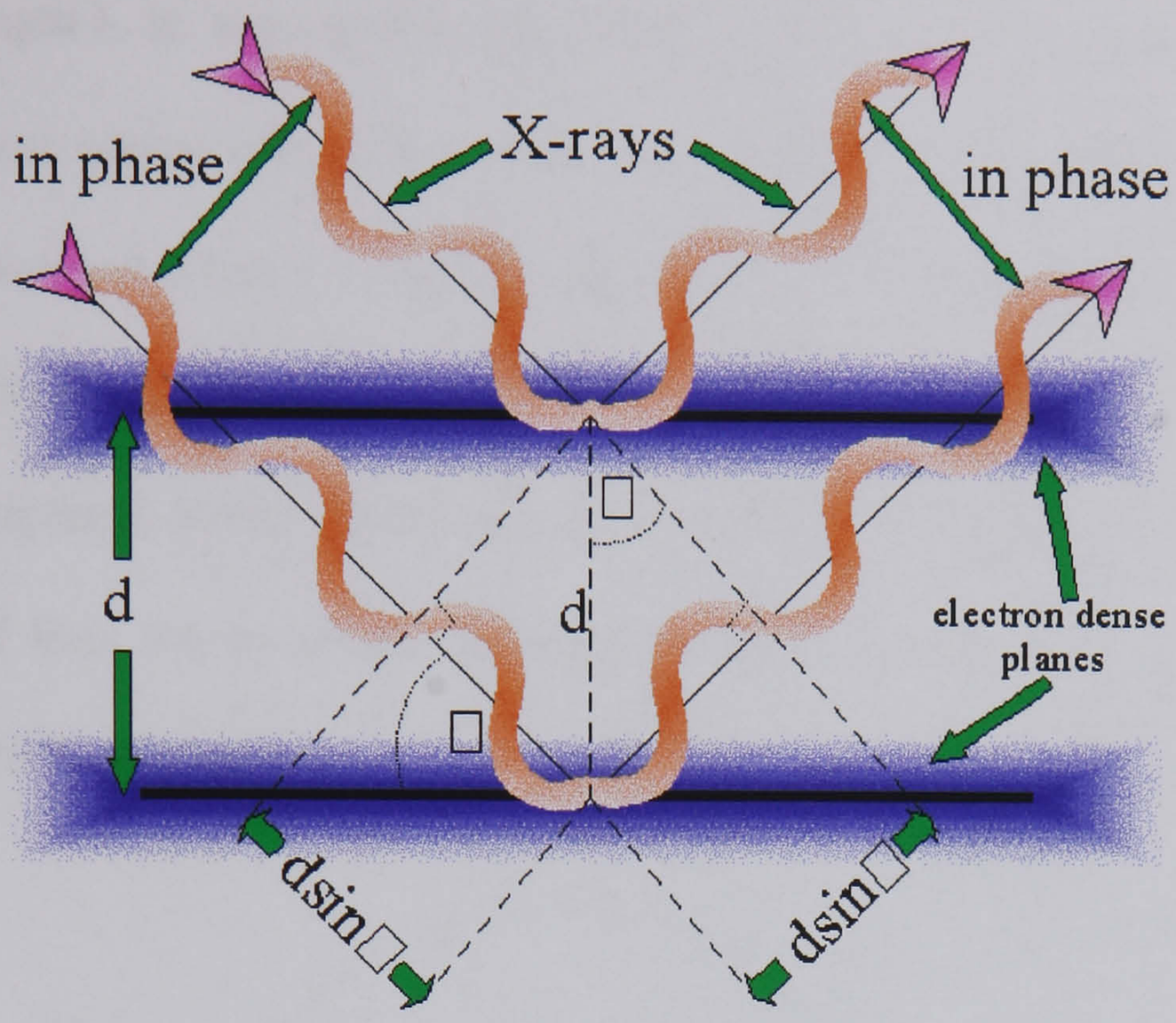


Figure 2.4 Parallel rays reflected from two successive planes.
In this case n = 1, however, this can apply any number of planes. The angle of diffraction is inversely porportional to the interplanar distance. d=interplanar distance, □□=diffraction angle.

The three-dimensional crystal lattice hence determines a three-dimensional lattice of scattered X-rays; this can be represented as the 'reciprocal lattice'. As the crystal rotates so does the reciprocal lattice.

The reciprocal lattice is a convenient abstract concept useful for constructing the directions of diffraction by a crystal. The reciprocal lattice spacings are related to the spacings of the reflections on the detector. However, the reciprocal lattice spacings are inversely related to the crystal lattice spacings : It is therefore possible to obtain unit cell dimensions from reflection spacings. The space group of the unit cell can be determined by looking for symmetry within the diffraction pattern and the absence of certain Bragg reflections.

Diffraction conditions are determined not only by the reciprocal lattice, but also by the radius of the 'Ewald sphere', which has a radius reciprocal to the X-

ray wavelength λ, ie. it is equal to 1/λ (Figure 2.5): Bragg's law is satisfied when a lattice point comes into contact with the sphere. Therefore, diffracted beams are only observed when a reciprocal lattice point is on, or passes through, the surface of the Ewald sphere.

Other reciprocal lattice points can be brought to diffraction by rotating the crystal, and thus the reciprocal lattice, causing the intersection of new lattice points on the Ewald sphere.

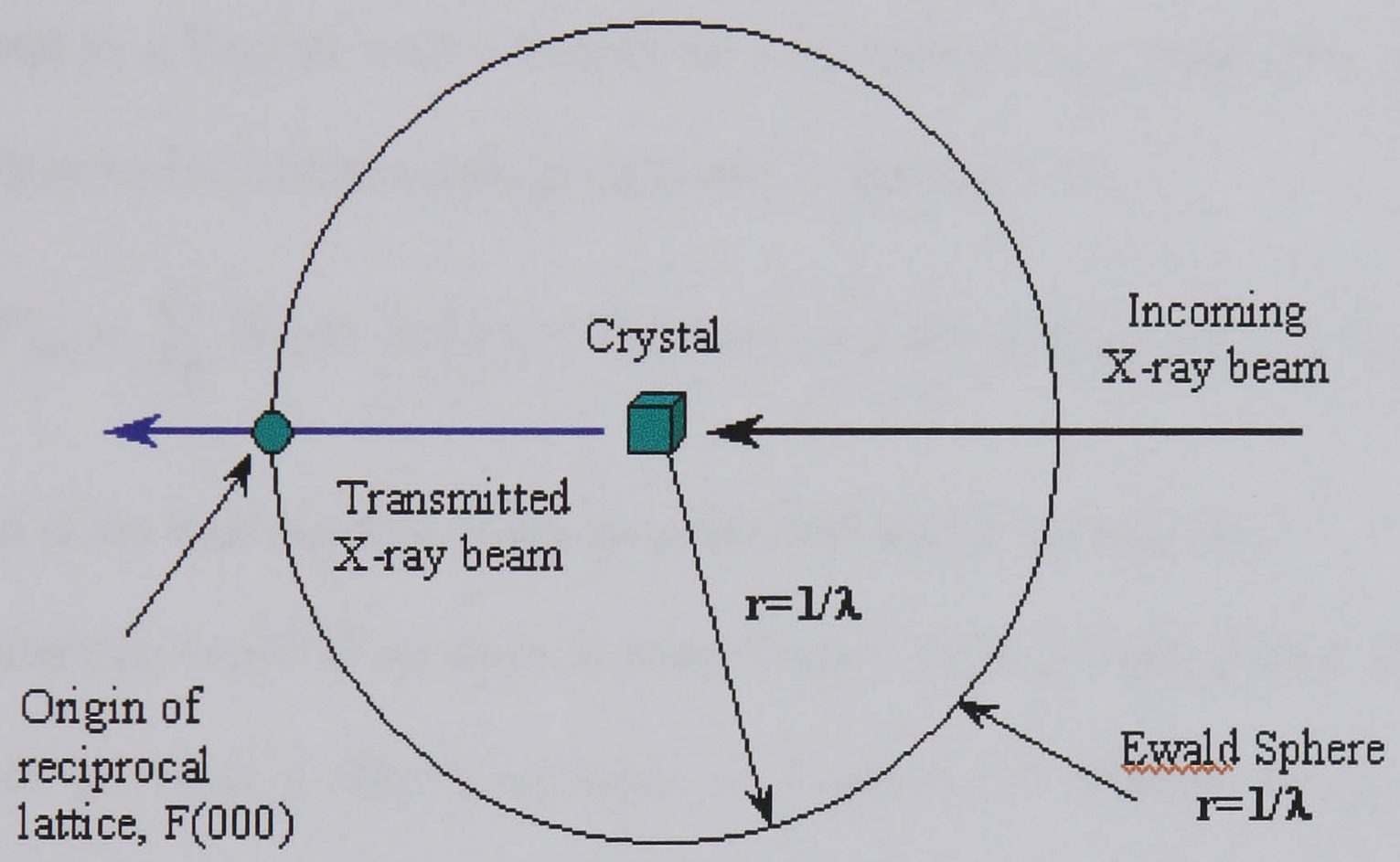


Figure 2.5 The generation of an X-ray diffraction pattern.

X-ray data collection results in a list of reflections: Each reflection is assigned an X-ray intensity and also indices *hkl*, that relate to its position in the diffraction pattern (*hkl* corresponding to the reciprocal lattice planes that produced the reflection).

The directions of the scattered waves are dependent upon the crystal lattice and the X-ray wavelength, and not upon the unit cell contents. The structural information about the contents of the unit cell is contained in the intensities of the reflections.

2.4 Molecular Structure from Diffraction Data

2.4.1 Calculating Electron Density

In order to ascertain a molecular structure from diffraction data the 'Fourier series' model of diffraction is applied. In this model, each atom or each infinitesimal volume of electron density is seen as the X-ray diffractor.

A diffracted beam, seen as a complicated wave, is the sum of diffractive contributions from all atoms (or infinitesimal electron density elements) within the unit cell. Therefore, each diffracted X-ray wave can be mathematically described as a Fourier series, known as a structure-factor, containing one term for each atom (or electron density element) in the unit cell:

$$F_{hkl} = \sum_{j=1}^{n} f_j[\cos 2\pi(hx_j + ky_j + lz_j) + i \sin 2\pi(hx_j + ky_j + lz_j)]$$

where n is the total number of atoms contained within the unit cell.

The electron cloud of an atom scatters X-rays. Term f_j is the atomic scattering factor of an atom j. The atom, seen as a sphere of electron density, has a scattering power proportional to the number of electrons in the atom; scattering is also dependent on atomic vibrations. The terms in parentheses describe a three-dimensional wave function: x_j, y_j and z_j represent fractional coordinates of atom j in the unit cell, ie. real space. The indices h, k and l, as well as corresponding to a specific reflection, are also the frequencies of the X-rays along the x-, y- and z-axes, respectively. The phase of a wave is only dependent on the position of the scattering atom in the unit cell: For instance, one atom j at the fractional coordinates (x_j, y_j, z_j) contributes to the hkl reflection with a diffracted wave of phase $2\pi(hx_j + ky_j + lz_j)$ (with respect to a reference wave).

During Fourier synthesis of the structure-factor equation the basic waveform used is the complex number [cos $2\pi(hx_j + ky_j + lz_j) + i \sin 2\pi(hx_j + ky_j + lz_j)$], where i is the imaginary number $\sqrt{-1}$. Having a complex number as the basic waveform enables the phase of a reflection to be determined (discussed later).

Using the number theory equality $\cos\theta + i \sin\theta = e^{i\theta}$, the three-dimensional wave function can be expressed as an exponential. Therefore the structure-factor for a specific reflection, F_{hkl}, can be re-written as:

$$F_{hkl} = \sum_{j=1}^{n} f_j e^{2\pi i(hx_j + ky_j + lz_j)}$$

The ultimate goal of X-ray crystallography is to calculate the average electron density ρ at every position x, y, z, ie. $\rho(x,y,z)$, in the unit cell into which the protein structure can be built. Firstly, in order to achieve this goal, the structure-factor is treated as the sum of contributions from infinitesimally small volume elements of electron density:

$$F_{hkl} = \iiint_{hkl} \rho(x,y,z) e^{2\pi i(hx + ky + lz)} \, dx \, dy \, dz$$

or

$$F_{hkl} = \int_V \rho(x,y,z) e^{2\pi i(hx + ky + lz)} \, dV$$

where V is the unit cell volume.

A function f(x) that is related to F(h) in the method shown below is defined as the Fourier transform of F(h) and similarly F(h) is the Fourier transform of f(x):

$$F(h) = \int_{-\infty}^{+\infty} f(x) e^{2\pi i(hx)} \, dx$$

and

$$f(x) = \int_{-\infty}^{+\infty} F(h) e^{-2\pi i(hx)} \, dh$$

The units of the variable h are reciprocal to those of x. The Fourier transform is thus a relationship between a structure and its diffraction pattern. Therefore, the Fourier transform of all the structure factors (which are in reciprocal space, with reciprocal length spacings in $Å^{-1}$), gives an image of the unit cell and its contents (in real space, with spacings in Å):

$$\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l} F_{hkl} e^{-2\pi i(hx + ky + lz)}$$

or

$$\rho(x,y,z) = \frac{1}{V}\sum_{hkl} F_{hkl} e^{-2\pi i(hx + ky + lz)}$$

ie. electron density is the Fourier transform of all the structure-factors in the diffraction pattern. (The electron density map is a contour map of $\rho(x,y,z)$).

The electron density equation above suggests that the quality of the crystal structure depends upon the number of reflections included in the summation of the Fourier synthesis. The Fourier synthesis consists of adding waves of varying frequency; the smaller the values of h, k and l, the greater the frequency. The high-frequency terms are reflections nearer the centre of the diffraction pattern (produced by crystal lattice planes with large interplanar spacings): These confer gross features to the three-dimensional electron density map. The detailed features of the map are conferred by the high-frequency terms near the edge of the diffraction pattern. The maximum resolution of a structure is taken to be the minimum interatomic spacing, d_{min}, included in the Fourier synthesis.

2.4.2 The Phase Problem

A structure-factor can be represented as a complex vector in two dimensions using a phase-vector (or Argand) diagram (Figure 2.6). In the Argand diagram,

the structure-factor is divided into a real and imaginary part where the amplitude is proportional to the length of the vector, $|A|$, and the phase (with respect to a reference wave) is represented by the angle α made with the real axis.

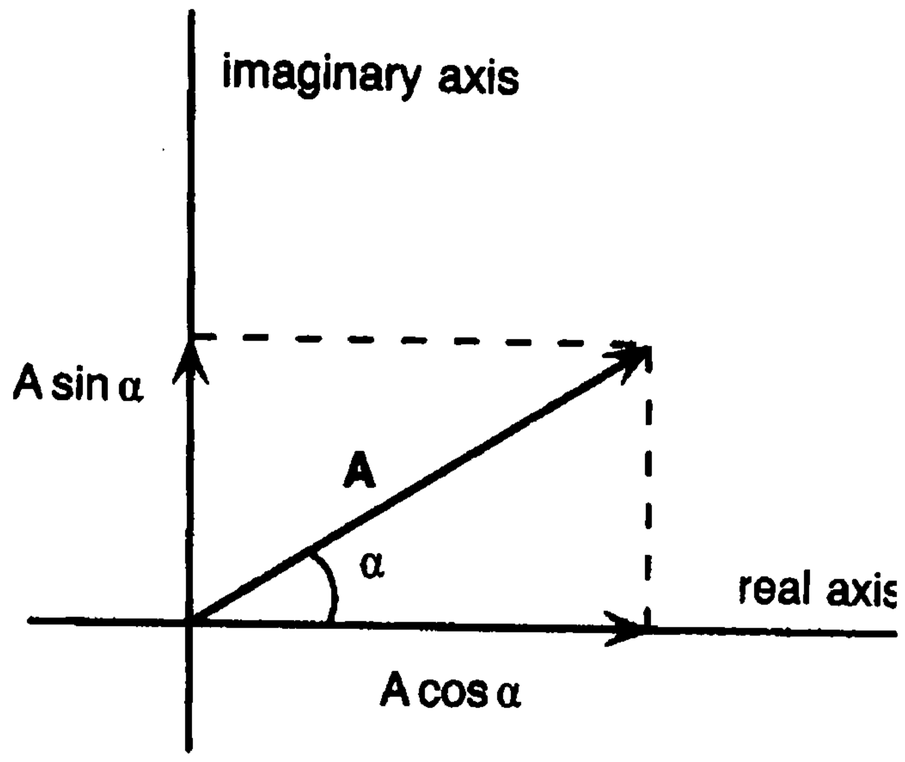


Figure 2.6 Argand diagram for a structure-factor.

Using simple vector arithmetic, the vector **A** may be expressed as:

$$\mathbf{A} = |\mathbf{A}|\cos\alpha + i(|\mathbf{A}|\sin\alpha)$$

or

$$\mathbf{A} = |\mathbf{A}|(\cos\alpha + i\sin\alpha)$$

where i is the imaginary number $\sqrt{-1}$.

Therefore, using the complex number equality $\cos\theta + i\sin\theta = e^{i\theta}$:

$$A = |\mathbf{A}|e^{i\alpha}$$

Thus, expressing the phase (in radians) of a reflection F_{hkl} as $\alpha_{hkl} = 2\pi\alpha'_{hkl}$, the electron density equation may be re-written as:

$$\rho(x,y,z) = \frac{1}{V}\sum_{hkl} |\mathbf{F}_{hkl}|\, e^{-2\pi i(hx + ky + lz - 2\pi\alpha'_{hkl})}$$

The amplitude $|\mathbf{F}_{hkl}|$ can be directly obtained from the intensity of the reflection (I_{hkl}) as $|\mathbf{F}_{hkl}| = (I_{hkl})^{1/2}$, subject to various polarisation and absorption correction factors. The frequencies of the the diffracted X-rays are given by the

indices of the reflection *hkl*. But, as the objective is to determine the atom positions within the unit cell, the phase information cannot be calculated; in effect the relative phases of the refections are lost in X-ray data collection.

In small molecule crystallography direct phasing methods can be used (e.g. Woolfson, 1997): The phase information can be revealed from statistical relationships in the intensity data of a crystal. Recently, direct phasing methods have been used utilised for a few small protein structures determined at atomic resolution. However, the overwhelming majority of crystallographic protein structures are not at atomic resolution and are too large for direct phasing methods; in these cases 'indirect' methods are used to obtain initial phases for each reflection.

2.4.3 The Molecular Replacement Method

2.4.3.1 Introduction

In protein crystallography initial phases for a first molecular model can be obtained if the structure of a similar or homologous protein is already known. For example, a previously determined crystal structure may be used as a 'search model' to obtain an initial electron density map for a mutant or inhibitor complex of the structure. Also, homologous proteins can be used to 'solve' (ie. obtain initial phases) new structures of interest: This is based on the observation that proteins which are homologous in amino acid sequence commonly have similar polypeptide chain folds.

Frequently, the known protein structure has a different crystalline arrangement to that of the unknown (usually new) structure. Therefore the search model has to be be transferred into the new target unit cell, this is achieved using the

molecular replacement method (Rossmann and Blow, 1962). If there is one molecule in the asymmetric unit, the transformation matrix relating the two structures requires six parameters: Three angles define the rotation required to bring the structures into the same orientation and three translation parameters superimpose the two structures.

A brute-force six-dimensional search for the required transformation matrix (with respect to an origin) is not viable as it is beyond the capacity of even the fastest modern computers. However, the search can be greatly simplified by splitting the operation into two three-dimensional searches (e.g. Driessen and Tickle, 1996): The rotational search is used to find the correct orientation of the search model and then the translational search is used to place the orientated model into the correct position within the target unit cell.

2.4.3.2 The Rotation Function

The rotational search is carried out by looking for agreement between the Patterson functions of the known and unknown models. The Patterson function is a Fourier transform of the set of squared structure-factor intensities without phases (h, k, l, $|\mathbf{F}_{hkl}|^2$):

$$P(u,v,w) = \frac{1}{V}\sum_{hkl} |\mathbf{F}_{hkl}|^2 e^{-2\pi i(hu + kv + lw)}$$

Patterson-space unit cells are of the same dimensions as real-space cells and the x, y and z dimensions of Patterson vector-space are called (u,v,w). The function does not produce an electron density map of the unit cell contents but rather a density map of the vectors between atoms of the unit cell (including vectors between the same pair of atoms but in opposite directions), superimposed and

emanating from the origin of the map. Self-vectors (vectors between equilvalent atoms) result in a peak at the origin. In a Patterson map there are $n(n - 1)$ vectors from a crystal containing n atoms in the unit cell.

The intramolecular Patterson function is independent of the position of the molecule within the unit cell, however, rotating the molecule alters the Patterson map (ie. changing the orientation of the structure changes the Patterson function).

For identical or closely related molecules in two different lattices, the orientation of the unknown structure is determined by superimposing the different Patterson maps by applying a 'cross-rotation function' that maximises the overlap between the maps. A search for molecular symmetry axes within a crystal involves comparing a Patterson map with itself in different orientations, and is known as the self-rotation function. Determination of any local symmetry axes by the self-rotation function will reduce the number of possible solutions to the cross-rotation function.

Most molecular replacement programs output a large number of possible rotation solutions. Each solution usually has a correlation coefficient, between the observed and calculated Patterson vectors, associated with it: The coefficient commonly reflects the product of the agreement at each position of the two Patterson maps. The product will only be large if both maps have high density coinciding at one position. The 'correct' rotation solution(s) will normally have a correlation coefficient well above the others.

2.4.3.3 The Translation Function

Translation searches are employed to determine the location of molecules (with known orientation) in a unit cell: The search is usually carried out by a trial and error process whereby the orientated model is placed at various positions; each time structure factor amplitudes are calculated ($|\mathbf{F}_{calc}|$) and compared with the observed structure factor amplitudes ($|\mathbf{F}_{obs}|$) by calculating an *R*-factor (described later) and a correlation factor. The correlation coefficient has the advantage over the *R*-factor in that the coefficient is insensitive to the intensity scale.

As with the rotation function, if more than one molecule occupies the asymmetric unit, the translation function of each molecule must be determined independently. In such cases, the subsequent translation solutions have even lower *R*-factors and greater correlation coefficients than the first solution.

2.4.3.4 Rigid body refinement

Some computer programs improve the molecular replacement solution by applying rigid body refinement; in this procedure a rigid geometry is assigned to the whole (or parts) of the model and attempts made to minimise the fit of $|\mathbf{F}_{calc}|$ to $|\mathbf{F}_{obs}|$.

2.4.4 The Multiple Isomorphous Replacement Method

2.4.4.1 Introduction

Molecular replacement requires the availability of a known homologous protein structure. If this does not exist, isomorphous replacement can be applied; this is a commonly used method for determining phase angles where nothing is known about the three-dimensional structure of the protein. This technique

requires the attachment of heavy atoms (atoms with high atomic number) to the protein molecules in a crystal (Kendrew *et al.*, 1958) Phases calculated from the positioned heavy atoms can then be used to obtain an initial electron density map.

Every atom in the unit cell contributes to all measured reflections, however, any atom's contribution is greatest to the reflection whose indices correspond to lattice planes that intersect the atom. So, if a few atoms were introduced into identical positions within all unit cells of a crystal, without any alteration in crystalline packing (isomorphism), the changes observed in the intensities of the diffraction pattern would be greater in some reflections than in others. Adding heavier atoms (larger scattering factors) to a crystal would increase the differences in the diffraction pattern which would then be large enough to measure.

2.4.4.2 Obtaining Phases from Derivative Data

If heavy atoms can be bound at a small number of well defined sites, with high occupancy, to the native protein to form a heavy atom derivative crystal, it may be possible to derive some phase information for the native dataset.

Assuming that $\mathbf{F}_P$ is the structure-factor of a reflection for the native protein, $\mathbf{F}_{PH}$ is the structure-factor for the corresponding reflection in the derivative crstal and that $\mathbf{F}_H$ is the structure-factor for the heavy atom, simple vector arithmetic shows that, for acentric reflections (ie. non-centric reflections):

$$\mathbf{F}_{PH} = \mathbf{F}_P + \mathbf{F}_H$$

and that $\mathbf{F}_H$ is the structure-factor for the heavy atom, simple vector arithmetic shows that, for acentric reflections (ie. non-centric reflections):

$$\mathbf{F}_{PH} = \mathbf{F}_P + \mathbf{F}_H$$

This is because all scattering atoms contribute to a reflection (Figure 2.7a). For centric reflections (where the phase angle α_P is 0 or π, if the origin of the system is placed in the centre of symmety) $|\mathbf{F}_P| = |\mathbf{F}_{PH}| \pm |\mathbf{F}_H|$ (Figure 2.7b).

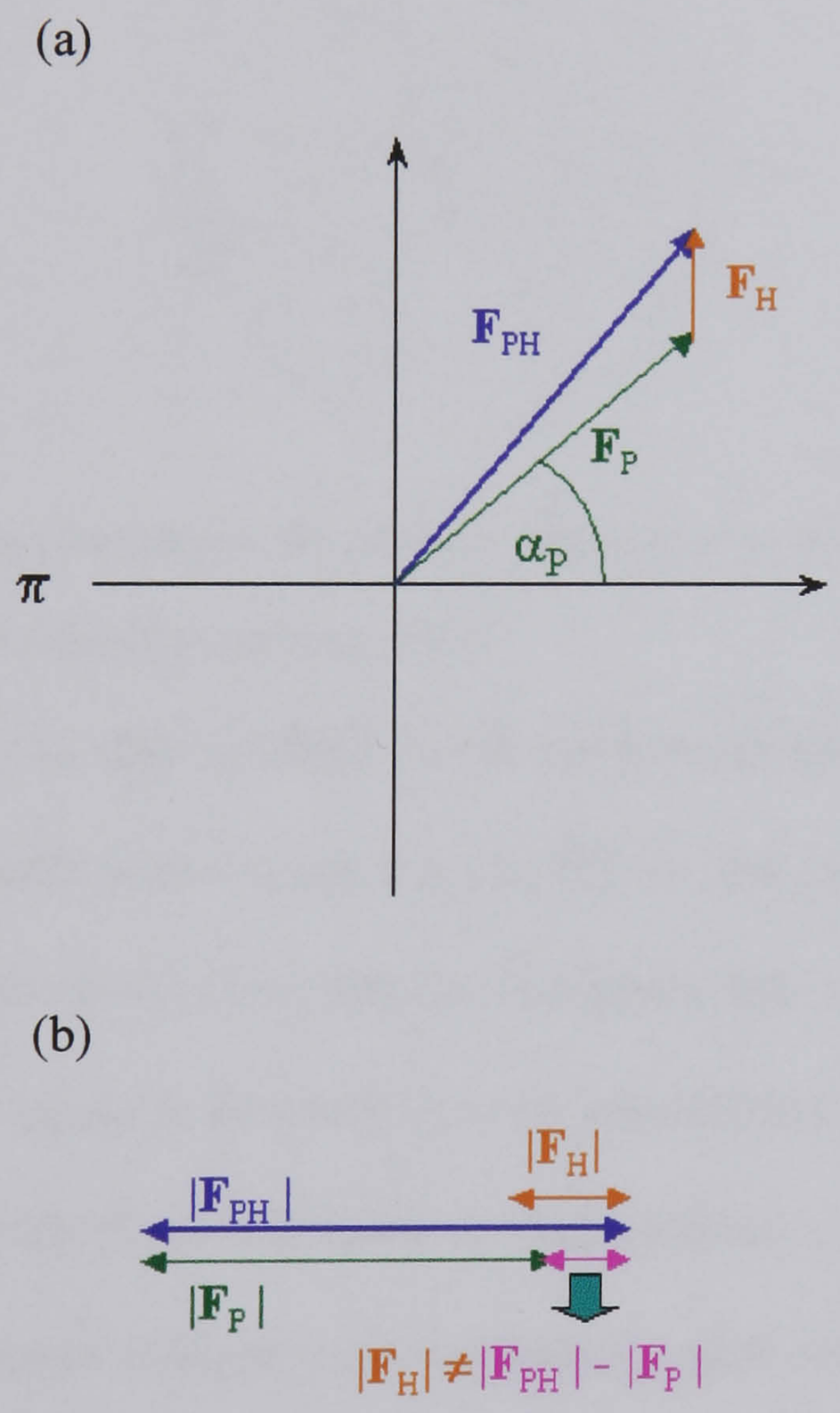


Figure 2.7 Structure factors in the isomorphous replacement phasing method for **(a)** acentric reflections **(b)** centric reflections.

$\mathbf{F}_H$ can be determined once the heavy atom(s) can be located. From the data collection experiments the magnitudes $|\mathbf{F}_P|$ and $|\mathbf{F}_{PH}|$ can be computed. However, in order to calculate an electron density map, the phase angle α_P of the

vector $\mathbf{F}_P$ has to be determined for each reflection. It follows from the previous equation that for acentric reflections:

$$\mathbf{F}_P = \mathbf{F}_{PH} - \mathbf{F}_H$$

To calculate α_P, the above vector equation is represented in a complex plane: The vector $-\mathbf{F}_H$ is placed at the origin and a circle of radius $|\mathbf{F}_{PH}|$ centered at the top of vector $-\mathbf{F}_H$. And a circle of radius $|\mathbf{F}_P|$ centered at the origin is drawn.

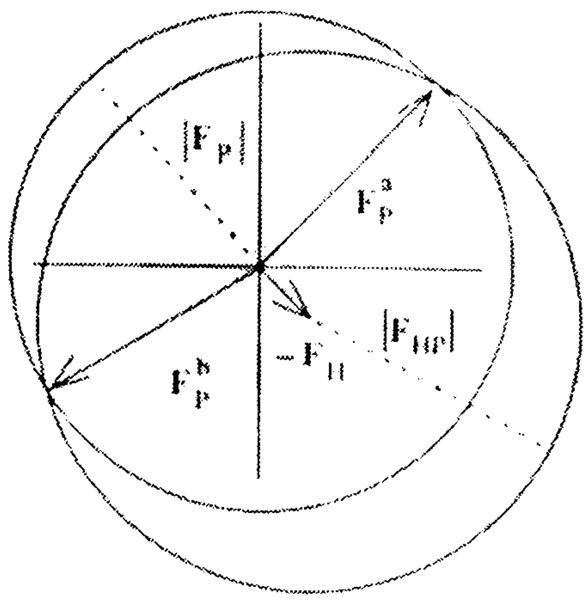
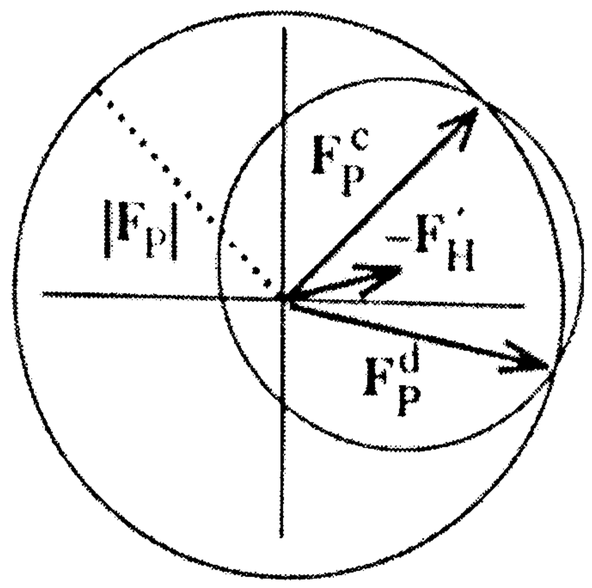


Figure 2.8 Determining phase angles for acentric reflections using vectors in the isomorphous replacement method. (Adapted from Rhodes, 1993).

The equation for $\mathbf{F}_P$ is only satisfied where the two circles intersect. Therefore the phase angles made with vectors $\mathbf{F}^a_P$ and $\mathbf{F}^b_P$ are the possible phases of the reflection. At this point it is clear that the best phase that can be obtained from the solutions is the mean in between the two possibilities. So, the phase error using the single isomorphous replacement (SIR) method can be quite large. In order to eliminate phase ambiguity a second heavy atom derivative ($\mathbf{F}'_H$) can be prepared and, provided that heavy atom is not at the same position, a unique solution for α_P can be found (Harker, 1956) (Figure 2.9).

(a)



(b)

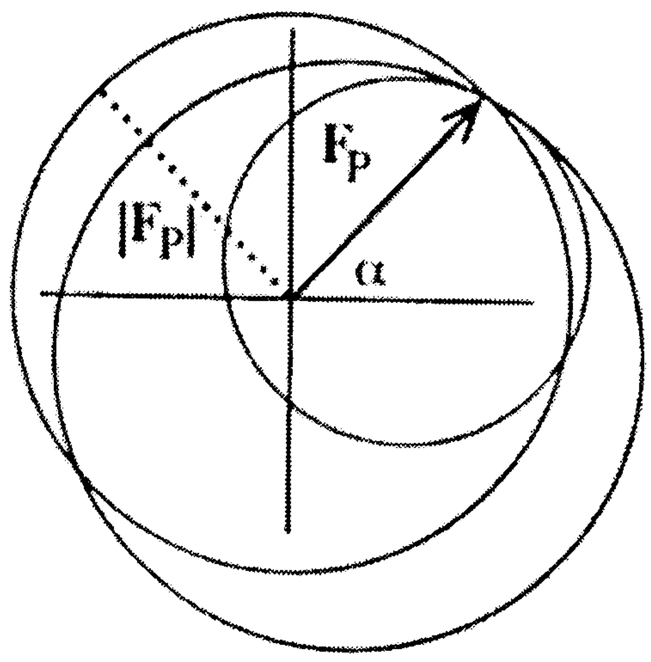


Figure 2.9 (a) A second derivative is prepared and again there are two possible phase angles for $\mathbf{F}_P$. **(b)** $\mathbf{F}^c_P$ corresponds with $\mathbf{F}^a_P$, so α is the the correct phase for $\mathbf{F}_P$.

In theory, the phase angle calculated for the reflection should be exact. But, this is based on several assumptions, including ideal isomorphism, precise measurement of X-ray intensities and knowledge of the exact heavy atom positions, none of which are perfectly met in practice. These imperfections may lead to the 'phasing circles' grazing one another, causing uncertainity in the phase angle (or the circles may not intersect at all). These uncertainies are measured in terms of 'phase probabilities': A probability density function will have a maximum at the most probable phase. Other derivatives may be

necessary to improve the quality of the phases. The method is therefore called multiple isomorphous replacement (MIR).

Data sets from heavy atom derivatives are often at a lower resolution when compared with the corresponding native data sets. Therefore, initially phase estimates are not available for native reflections at a higher resolution than that of the best heavy atom derivative. Generally, for useful phase information to be extracted a derivative crystal has to diffract to at least 4.5Å.

For perfect isomorphism the crystalline packing must be identical. The intensity differences between the native and derivative data sets are then exclusively due to the attached heavy atoms. Perfect isomorphism hardly ever occurs. Errors due to nonisomorphism are commonly one of the main obstacles in MIR (as it causes difficulties during heavy atom position refinement). Nonisomorphism can manifest itself as a slight change in protein conformation as a consequence of heavy atom binding, but it usually presents itself as a change in the unit cell dimensions. Normally, in practice changes in cell lengths relative to the native crystal of up to ~ 0.5% can be tolerated.

There are several parameters for assessing the overall quality of phase information obtained from a derivative dataset (Wolf *et al.*, 1991). These include the occupancy of the heavy atoms, R_{Cullis}, phasing power and R_{iso}.

R_{Cullis} for centric reflections is expressed as:

$$R_{Cullis} = \frac{\sum_{hkl} \| F_{PH} \pm F_P | - F_H(\text{calc}) |}{\sum_{hkl} | F_{PH} \pm F_P |}$$

In practice the obsevered amplitude $|\mathbf{F}_{PH}|$ will be too short or too long to exactly reach the endpoint of $\mathbf{F}_H$ (Figure 2.7). The difference is called the 'lack of closure error', so in essence the R_{Cullis} is:

<lack of closure> / <isomorphous difference>

An R_{Cullis} for centric reflections (expressed as a function of resolution) less than 0.6 is considered to be excellent and anything less than 0.9 is usable.

The phasing power is the measure of the quality of a derivative and it is expressed as:

$$<|\mathbf{F}_H|> / <\text{lack of closure}>$$

Phasing power values greater than 1.5 are considered excellent and anything more than 0.5 is usable.

R_{iso} is the isomorphous difference ($\sum_{hkl}|F_{PH} \pm F_P|$) ; for a good derivative it should normally be between ~ 0.15 and 0.25 (ideally, if no heavy atoms were bound then the R_{iso} would be zero). If the nonisomorphism between the native and derivative crystals is appreciable, the R_{iso} will appear larger than it really is.

2.4.4.3 Locating heavy atom positions

The phase and magnitude of $\mathbf{F}_H$ can be calculated easily if the positions of the heavy atoms within the unit cell are known.

Patterson functions are formulated as squares of the number of electrons of the scattering atoms, therefore Patterson maps of heavy atom derivative crystals are dominated by vectors between heavy atoms. However, these protein derivative Patterson maps are usually not interpretable because the number of peaks in a

Patterson map is also related to the square of the number of atoms. To overcome this problem, a difference Patterson function is used to locate the heavy atoms.

A difference Patterson map $(|\mathbf{F}_{PH}| - |\mathbf{F}_{P}|)^2$ is calculated with the amplitudes of a derivative crystal and the amplitudes of an isomorphous native crystal:

$$\Delta P(u,v,w) = \frac{1}{V} \sum_{hkl} \Delta \mathbf{F}^2_{hkl} e^{-2\pi i(hu + kv + lw)}$$

where $\Delta \mathbf{F}^2_{hkl} = (|\mathbf{F}_{PH}| - |\mathbf{F}_{P}|)^2$.

Therefore, the scatter of the light protein atoms is removed and the difference map shows just the vectors between heavy atoms (and noise).

Symmetry within the unit cell can be used to locate heavy atom positions in the difference Patterson map. For example, if a protein crystallises in the $P2_1$ space group (a two-fold screw axis along **c**) an atom at (x,y,z) will have an identical symmetry-related atom at $(-x, -y, ½ + z)$. Therefore the vectors between such atoms in a difference Patterson map will all lie at $(2x, 2y, -½)$. As these are in fractional unit cell coordinates, $(u,v,w) \equiv (2x, 2y, ½)$. All heavy atom vectors lie in the Patterson unit cell plane $w = ½$. These planes, that contain difference Patterson vectors between symmetry-related atoms, are known as Harker sections. If heavy atoms are bound such that they are attached at equivalent positions throughout the crystal, the heavy atoms peaks can be found on the Harker sections. Once a heavy atom is found its position can be refined using least-squares or maximum-likelihood methods (described later) and then subsequent heavy atoms sought.

2.4.4.4 Preparing heavy atom derivative crystals

The search for heavy atom derivatives is still essentially a trial-and-error process. The preferred derivative preparation method involves soaking the protein crystal in a solution of the heavy atom reagent, this is the technique presented in the thesis. Other derivative procedures such as chemical or genetic modification of the protein and xenon atom attachment (Drenth, 1995) exist; however, these methods are as yet not as successful as soaking heavy atoms into a crystal.

(Cocrystallisation is not often employed because the crystals tend not to grow or grow nonisomorphously).

A considerable volume of the protein crystal is occupied by water (Matthews, 1968), providing solvent channels and pores where extra molecules may enter and bind to the protein without disturbing the crystal lattice. However, some sites will be more or less accessible than others. Therefore, multiple sites at less than full occupancy are common.

The heavy atom soaking solution is identical in composition to the mother liquor, apart from the added reagent and often a slight increase in precipitant concentration. After soaking, if the heavy atom occupancy is not very high the reagent concentration can sometimes be increased, but the chances of nonisomorphism or even crystal degradation are greater. The soaking time varies between minutes and months; it depends on the time required for the reaction to reach equilibrium, this is determined by a number of factors. For

example, the diffusion rate is dependent on the crystal's solvent content and on the size of the heavy atom compound.

Although the search for suitable heavy atom reagents is an empirical process, the biochemical properties of the protein, and the chemical properties of the buffers & precipitants, can be employed in the search (Blundell and Johnson, 1976). For instance, histidine residues often act as ligands to heavy atoms such as uranium or samarium. The pH should not be too low as it is the neutral histidine side chain that acts as the ligand: However, if the pH is too alkali some heavy atom salts, such as those containing Sm^{3+}, are easily hydrolysed. Platinum and mercury compounds tend to be more reactive at lower pH values; platinum compounds commonly binding to the the sulphur atom in methionine, whereas mercury compounds tend to associate with free sulphide groups. Still, it is often the case that many reagents are tried before a potential derivative is found.

2.4.5 The Multiwavelength Anomalous Dispersion Method

2.4.5.1 Introduction

Like MIR, the method of multiwavelength anomalous dispersion (MAD) is a de novo phasing approach. But, unlike MIR, the MAD technique takes advantage of anomalous diffraction in determining the phase angles (Hendrickson, 1991).

Anomalous scattering is a phenomenon which influences the intensities of X-ray reflections and causes a breakdown of Friedel's Law so that $I_{hkl} \neq I_{-h-k-l}$. The difference in intensities of Friedel pairs in an anomalous scattering data set can establish their phases in the nonanomalous scattering data set.

2.4.5.2 Anomalous Scattering

All elements absorb as well as emit radiation. There is a sudden drop in absorption just below the characteristic emission wavelengths (Figure 2.2). This is known as the 'absorption edge' of the element. An element exhibits anomalous scattering when the incident X-ray wavelength is close to the absorption edge of the element.

Thus far, X-ray diffraction has been treated as scattering from a free electron of an isolated atom. However, a basic model of anomalous diffraction can be described if the isolated scattering atom's electron is related to bound electronic states of the atom (James, 1948).

If an incident photon has relatively low energy it is either scattered or not, but it cannot be absorbed as it has insufficient energy to cause electronic transitions. Therefore, as the photon scatters with no phase delay, only the normal atomic scattering coefficient f° can be used to describe the scattering.

On the other hand, if the incident photon has a high enough energy, some photons will be scattered normally, some absorbed and re-emitted at lower energy (fluorescence), and some photons will be absorbed and immediately re-emitted at the same energy: In the latter case, the photon is retarded compared to a normally scattered photon, thus the scattered photon can be seen as gaining an imaginary component f'' to its phase.

This effect can be measured as a function of X-ray energy by observing the sharp increase in absorption or fluorescence: The imaginary anomalous scattering component f'' is proportional to these measurable quantities. An

anomalous scattering profile for isolated atoms of a typical element is shown in Figure 2.10.

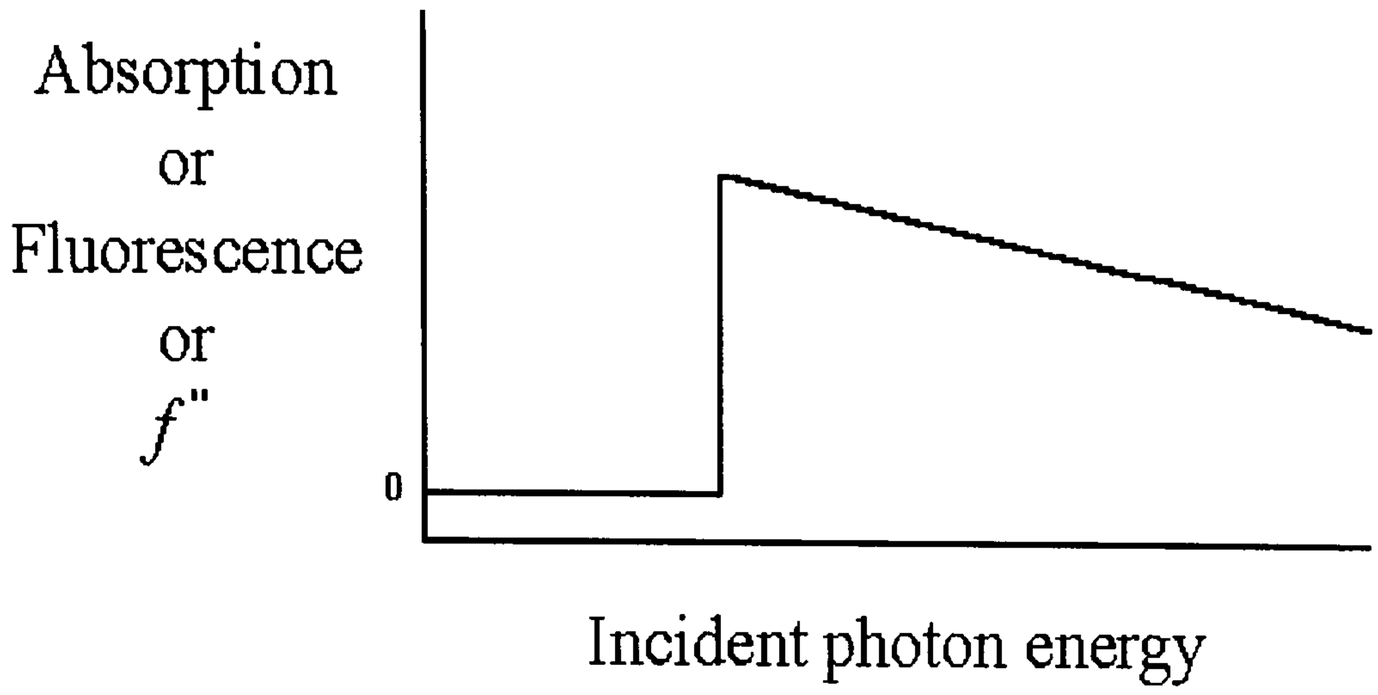


Figure 2.10 Profile of the imaginary anomalous scattering component f'' of isolated elemental as a function of X-ray energy.

The real component of the anomalous scattering f' is related to f'' by the Kramers-Kronig transformation (Kronig and Kramers, 1928). The total scattering factor is then

$$f = f^{\circ} + f' + if''$$

2.4.5.3 Effects of anomalous scattering

Using the structure-factor equation it is possible to show that a specific reflection F_{hkl} and its 'opposite' F_{-h-k-l} will have the same magnitude and opposite phases (ie. $I_{hkl} = I_{-h-k-l}$ and $\alpha_{hkl} = -\alpha_{-h-k-l}$), this is known as Friedel's Law. In the presence of anomalous contributions, the magnitudes of Friedel pairs are no longer identical and the phases are no longer centrosymmetric around the complex plane origin. Differences observed between Friedel pairs at the same wavelength are know as 'Bijvoet differences', whereas differences observed in the same F_{hkl} at different wavelengths are called 'dispersive differences'. Both the dispersive and Bijvoet differences are used for phasing in a multiwavelength

experiment. The changes observed in intensity are very small, usually only a few percent between reflection pairs.

Centric reflection pairs F_{hkl} and F_{-h-k-l} are related through the space group's point symmetry (Laue symmetry) and are not Friedel pairs. Centric reflections can be used to determine data quality statistics (e.g. R_{merge}) in the absence of anomalous contributions. A comparison of centric reflection statistics to the merged Friedel pairs (which should merge significantly higher) indicates whether a useable anomalous signal is present.

A structure-factor $\mathbf{F}_{HP}{}^{\lambda 1+}$ (the H indicating an anomalous scatter) measured at a wavelength λ_1, where anomalous diffraction does not occur can be represented in a complex plane as shown in Figure 2.11. If the same structure factor is measured at a wavelength λ_2 where anomalous diffraction occurs, the structure-factor $\mathbf{F}_{HP}{}^{\lambda 2+}$ will now have contributions from the normal scattering component $\mathbf{F}_{HP}{}^{\lambda 1+}$ and an anomalous scattering component, the latter consisting of two perpendicular vectors one of which is real ($\Delta\mathbf{F}_r^+$) and the other imaginary ($\Delta\mathbf{F}_i^+$).

At the wavelength λ_1 where there is no anomalous scattering, there will be a Friedel mate $\mathbf{F}_{HP}{}^{\lambda 1-}$ which is a reflection in the real axis of $\mathbf{F}_{HP}{}^{\lambda 1+}$. However, at wavelength λ_2 anomalous diffraction occurs and thus Friedel's Law is broken: $\Delta\mathbf{F}_r^-$ is the real axis relection of $\Delta\mathbf{F}_r^+$, but it can be proved that the imaginary anomalous contribution $\Delta\mathbf{F}_i^-$ is the inverted reflection of $\Delta\mathbf{F}_i^+$. Therefore $\mathbf{F}_{HP}{}^{\lambda 2+}$ and $\mathbf{F}_{HP}{}^{\lambda 2-}$ are not Friedel pairs.

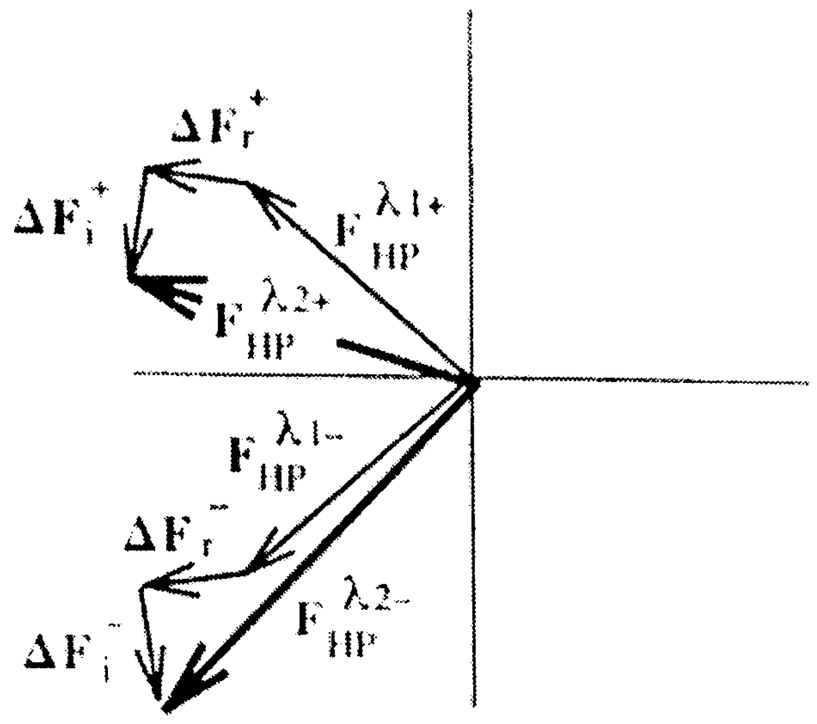


Figure 2.11 Structure-factors under anomalous scattering. (Adapted from Rhodes, 1993).

2.4.5.4 Obtaining Phases from anomalous scattering data

The phase angle for each structure-factor has to be determined. Considering the structure-factor $\mathbf{F}_{HP}^{\lambda 1+}$ it is known from Figure 2.11 that:-

$$\mathbf{F}_{HP}^{\lambda 1+} = \mathbf{F}_{HP}^{\lambda 2+} - \Delta\mathbf{F}_r^{+} - \Delta\mathbf{F}_i^{+}$$

The phases of the anomalous components (real and imaginary) depend only upon the location of the anomalous scatters within the unit cell, these positions can be found using Patterson methods (described later). Also, the magnitudes of $\Delta\mathbf{F}_r$ and $\Delta\mathbf{F}_i$ are fairly constant for a given element and are independent of the reflection angle, the magnitudes of $\mathbf{F}_{HP}^{\lambda 1+}$ and $\mathbf{F}_{HP}^{\lambda 2+}$ can be directly determined. As with the MIR method, the above equation can be solved for $\mathbf{F}_{HP}^{\lambda 1+}$ geometrically using a Harker diagram (Figure 2.12).

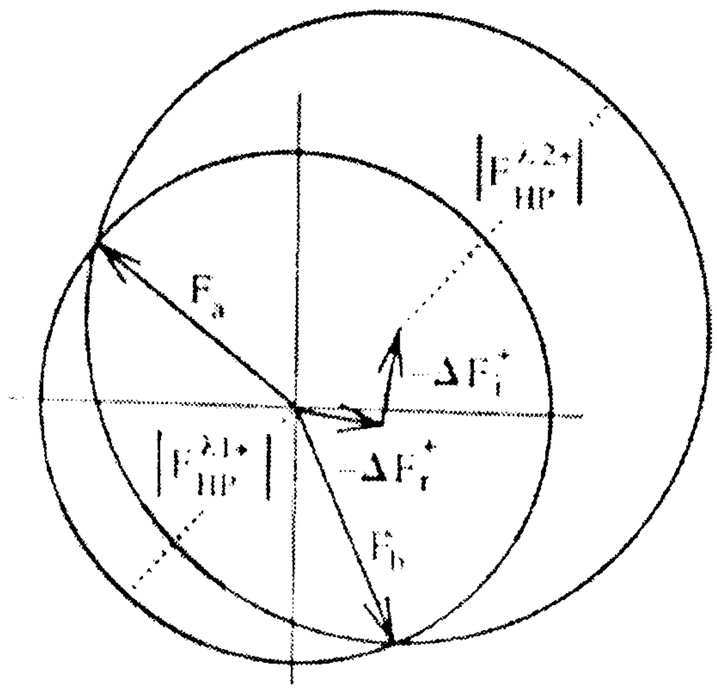


Figure 2.12 Harker diagram determining the phase of a structure-factor.

There is some phase ambiguity, the phases of both $\mathbf{F}_a$ and $\mathbf{F}_b$ both solve the equation. However, the structure-factors Friedel mate helps in the correct phase determination. A second equation can be derived for $\mathbf{F}_{HP}{}^{\lambda 1+}$ by reflecting $\mathbf{F}_{HP}{}^{\lambda 2-}$ and its components across the real axis (Figure 2.13a). The magnitude of $\mathbf{F}_{HP}{}^{\lambda 2-}$ is still the same, $\Delta\mathbf{F}_i^+ = \Delta\mathbf{F}_i^-$ and $\Delta\mathbf{F}_r^+ = \Delta\mathbf{F}_r^-$, so:

$$\mathbf{F}_{HP}{}^{\lambda 1+} = \left| \mathbf{F}_{HP}{}^{\lambda 2-} \right| - \Delta\mathbf{F}_r^+ + \Delta\mathbf{F}_i^+$$

Again, the above equation can be solved geometrically (Figure 2.13b). The phase ambiguity has been removed as one of the solutions to the equation corresponds with $\mathbf{F}_a$.

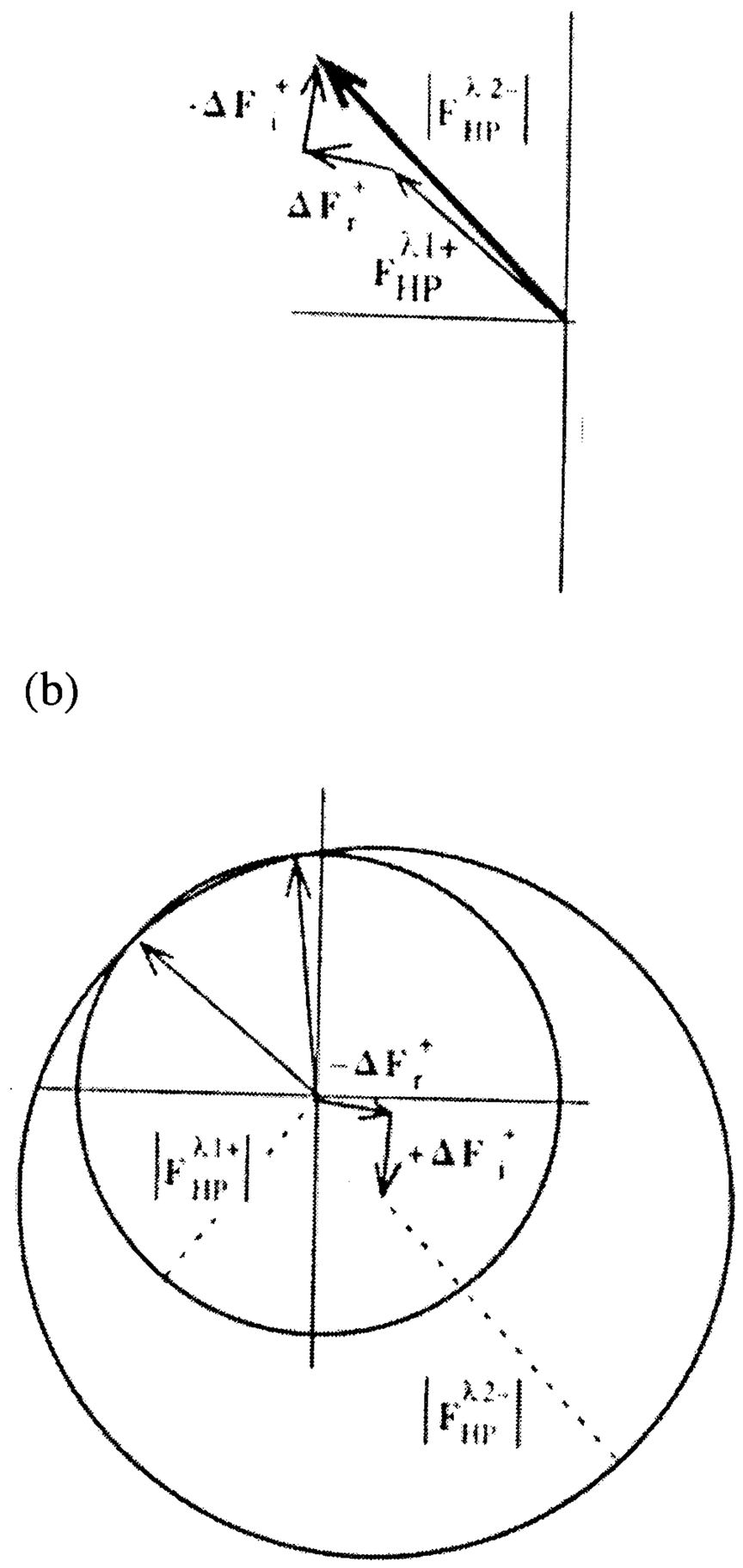


Figure 2.13 (a) Reflection of a structure-factor and its components across the real axis **(b)** Removal of phase ambiguity. (Adapted from Rhodes, 1993).

(As with MIR programs, MAD computer programs calculate phase numerically rather than geometrically by obtaining solutions to a series of simultaneous equations).

One of the advantages of using MAD over MIR is that there are no nonisomorphism problems associated with anomalous scattering, this is because all the diffraction data required for MAD phasing can be collected from a single crystal. However, an isomorphous replacement signal is generally a lot stronger

than an anomalous signal. The phasing methods may also be combined, for example if the multiple isomorphous method includes anomalous scattering information, it is called the MIRAS method (or SIRAS if a single derivative is used). MIRAS is particularly useful in improving the protein phase angles of MIR or MAD data.

Generally, for useful phase information to be extracted from anomalous data sets the crystal has to diffract to at least 3.5Å (Hendrickson, 1991). The success of MAD phasing also depends on data quality and completeness, and special techniques for collecting Friedel opposites, such as the inverse beam geometry method; this method reduces the effects of radiation damage, etc. by measuring Bijvoet pairs close together in time by exploiting centrosymmetry in the diffraction pattern.

2.4.5.5 Locating anomalous scatters

If the anomalous scattering atoms within the unit cell can be located then the corresponding phase contributions (real and imaginary) can be calculated. However, if the anomalous scatters are not heavy atoms which have been soaked into a crystal, a difference Patterson map cannot be used to locate them. In such cases, one of the functions that could be calculated is a 'Bijvoet difference Patterson' map with coefficients $(|\mathbf{F}_{hkl}| - |\mathbf{F}_{-h-k-l}|)^2$.

$$\Delta P(u,v,w) = \frac{1}{V}\sum_{hkl} \Delta\mathbf{F}^2 e^{-2\pi i(hu + kv + lw)}$$

where $\Delta\mathbf{F}^2 = (|\mathbf{F}_{hkl}| - |\mathbf{F}_{-h-k-l}|)^2$.

To increase the signal-to-noise ratio the $|\mathbf{F}_{hkl}|$ and $|\mathbf{F}_{-h-k-l}|$ values chosen would be from the wavelength that had the largest value of f''.

2.4.5.6 Choosing wavelengths for MAD data collection

The Cromer-Liberman theoretical values for scattering factors are not very accurate near an absorption edge. This is because the anomalous scattering atom is not isolated and is affected by the chemical environment (Smith, 1991). Therefore for each MAD experiment, X-ray absorption spectra are measured by fluorescence detection and used to directly determine the profile of the imaginary anomalous scattering component f'' as a function of wavelength (Hendrickson *et al.*, 1988). The f' values can then be derived from the corresponding f'' spectra using the Kramers-Kronig relationship.

Once the actual scattering factors f'' and f' have been determined in the energy range of the sample crystal, the exact X-ray wavelengths have to be chosen for the MAD data collection. For a MAD experiment at a single aborption edge, at least 2 wavelengths are required; however, it is best to use three or four so that the phase can be over-determined. The first wavelength to choose is the one with the largest signal, ie. maximal f''; this is indicated as λ_1 in Figure 2.14. The second wavelength often chosen is λ_2, the point of inflexion in f', with the maximal $|f'|$. The third and fourth wavelengths are at points remote from the absorption edge.

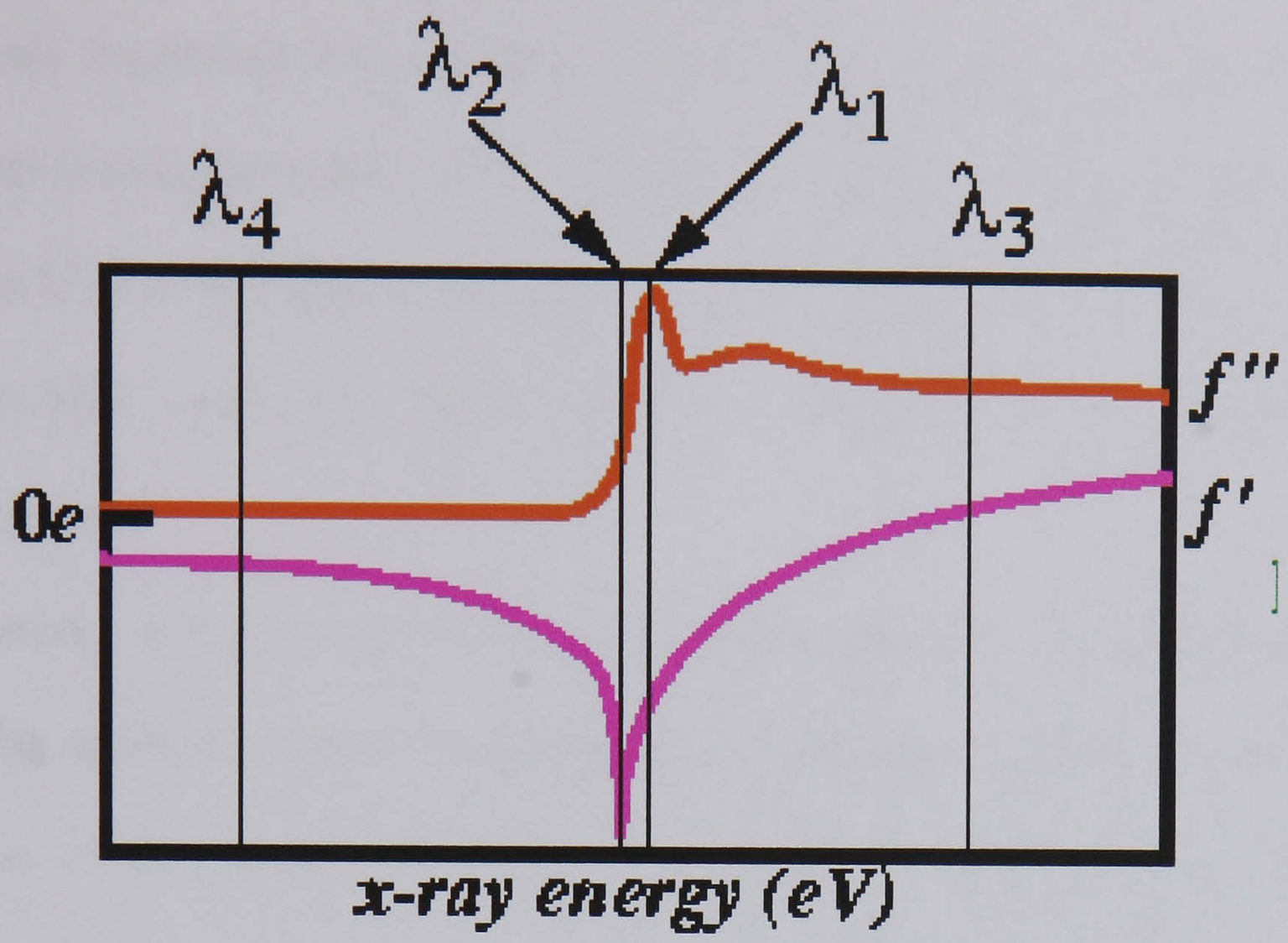


Figure 2.14 Schematic plot of experimental values for f' and f'' as a function of X-ray energy. (NB. 1KeV = 12.398 / λ in Å).

MAD experiments are possible using 'in-house' X-ray sources, but there are severe limitations due to the sporadic availability of characteristic emissions and the weakness of the Bremsstrahlung. So, almost all MAD phasing experiments are performed at radiation-dedicated synchrotrons. This is because synchrotrons are not only expectionally brighter X-ray sources but also any suitable wavelength in the spectral range can be selected with a monochromator. In addition, the selected radiation has a very narrow wavelength bandwidth; this is highly desirable for MAD phasing as the f'' and f' contributions to the atomic scattering factor change considerably with wavelengths near the X-ray absorption edge(s) of the element.

2.4.5.7 Incorporation of anomalous scatters

The protein should contain an element that gives a sufficiently strong anomalous signal. However, the elemental absorption edge(s) must be within the accessible X-ray spectrum to be useful. Using synchrotron radiation, the

accessible window is from ~0.3 to ~3.0Å. This range includes the K edges (electronic transitions in the K-shell) from Ca (Z = 20) to Ce (Z = 58); and the L_{III} edges (electronic transitions associated with the six 2*p* electrons) from Sb (Z = 51) to U (Z = 92) (Hendrickson, 1991). Several of the heavy atom elements used in MIR and a few lighter elements can in theory be used in MAD experiments.

Frequently, new protein structures have been determined from the anomalous scattering of Se (Z = 34) in selenomethionine. Selenomethionine (Se-Met) is an analogue of the amino acid methionine, where the sulphur atom has been replaced with a selenium atom. The X-ray absorption edge of selenium (λ = 0.98Å) is easily accessible with synchrotron radiation. Often, selenomethionine can replace methionine residues fully without perturbing the protein's structure or function (Hendrickson *et al.*, 1990). Incorporation of Se-Met into a protein does not perturb the macromolecular structure or function. It has been suggested that approximately one Se atom per 150 amino acids in a protein is sufficient for a successful application of MAD (Hendrickson *et al.*, 1990), however this is only a rule of thumb and it depends strongly on the quality of the diffraction data.

There are several ways in which selenium can be introduced into the protein. There were two methods employed in this thesis (described in detail in chapter 5): The first method involved growing the bacterial host strain on a selenomethionine substrate instead of a methionine-containing substrate, and blocking the host's methionine biosynthetic pathway using an excess of six amino acids. The second procedure used recombinant techniques in which the

protein was expressed in a methionine auxotroph and grown with selenomethionine as a nutrient. In both cases, the degree to which selenomethionine substituted methionine in the target protein was assessed using mass spectrometry.

2.5 Protein Structure Refinement

2.5.1 The Initial Model

After obtaining the initial experimental phases from molecular replacement, MIR, or MAD an initial model of the protein structure can be built into the first electron density maps. As the initial phases obtained are only estimates of the true phases, the model will contain errors and is likely to have poorly defined regions. However, the overall fold/conformation of the protein is likely to be correct.

2.5.2 Structure Refinement

2.5.2.1 Introduction

Crystallographic refinement aims to improve the fit of the model to the observed diffraction data. This is achieved by fitting $|F_{calc}|$ to be as near as possible to $|F_{obs}|$. Three methods of refinement were used, maximum likelihood, least squares, and molecular dynamics. Refinement has a limited radius of convergence, so parts of the model which are in error will not refine to their correct positions.

2.5.2.1 Refinement of protein structures

In order to define the atomic positions the coordinates x, y, and z are required. However, it is common to use the temperature (or B-) factor as a parameter. Therefore, for each atomic position four parameters have to be refined. This presents a problem in protein refinement, that of poor parameter:observation ratio.Therefore protein structures are almost always refined with known chemical information as a restraints. Not only do covalent (bond lengths and angles, Engh and Huber, 1991) and non covalent (hydrogen bonds and Van der Waals interactions) restraints prevent chemically incorrect structures, they also allow the parameters to be fitted, and therefore increase the observation:parameter ratio.

2.5.2.2 Maximum likelihood refinement

Maximum likelihood refinement is based on probability statistics. Using the constraints imposed by the experimental data, maximum likelihood refinement considers the likelihood of a particular reflection being observed and produces the model most likely to have given rise to that data. It also takes into account the completeness of the data and errors in the coordinates, and thus tends to produce more complete density than least squares refinement.

2.5.2.3 Least squares refinement

Least squares refinement attempts to refine structures usually in reciprocal space, minimising the difference between the observed and calculated electron densities. It involves adjusting the positions of the atoms in the asymmetric unit until the sum of the squared differences between $|F_{calc}|$ and $|F_{obs}|$ is minimised.

2.5.2.4 Refinement by simulated annealing

Simulated annealing (or molecular dynamics) refinement (Brünger *et al.*, 1987) is the most popular method of protein structure refinement, by subjecting the model to a large increase in temperature (usually around 3000K). The major advantage of using these high and obviously biologically unrealistic temperatures is that the model can transverse local minima and the radius of convergence is significantly increased when compared to techniques such as the classical gradient descent (conjugate-gradient and other least-squares methods). Simulated annealing was the refinement method used most extensively in the work presented in this thesis.

Simulated annealing searches for the conformation of the structure with the lowest energy, with the agreement with the diffraction data as part of the function. This minimisation function can be represented as :

$$E = E_{\text{chem}} + w_{\text{xray}} E_{\text{xray}}$$

where E_{chem} gives information about chemical interactions, E_{xray} describes the difference between the observed and calculated structure factor amplitudes in a form that is similar to that for least squares equation :

$$E_{\text{xray}} = \sum_{hkl} \left[F_{obs}(hkl) - K \left| F_{calc}(hkl) \right| \right]^2$$

where K is a scale factor. w_{xray} is a weighting factor which that balances the terms E_{chem} and E_{xray}. If w_{xray} is too large the refinement is weighted towards the diffraction data and the structure will show unrealistic deviations from ideal

geometry. If w_{xray} is too small the refined structure will not satisfy the diffraction data.

2.5.2.5 Rebuilding the model

The model was rebuilt in O (Jones et al., 1991), using difference maps. Rebuilding was based on the $|3F_{obs} - 2F_{calc}|$ and $|2F_{obs} - F_{calc}|$ maps, taking into consideration the $|F_{obs} - F_{calc}|$ and $|-F_{obs} - F_{calc}|$ maps. Side chain positions, and short areas of backbone were rebuilt accounting for the positions of surrounding residues, symmetry equilvalents, etc.

2.5.2.6 Monitoring the progress of refinement

The goal of refinement is to produce a chemically realistic model that satisfies the diffraction data. The overall progress of crystallographic refinement is traditionally monitored by the R-factor R_{cryst} :

$$R_{cryst} = \frac{\sum_{hkl} w_{hkl} \left(|F_{obs}(hkl)| - k|F_{calc}(hkl)| \right)}{\sum_{hkl} |F_{obs}(hkl)|}$$

where w_{hkl} are weights and k is a scale factor.

The R_{cryst} should be viewed with some caution because an acceptable value may be obtained through over-fitting the data, giving an incorrect structure with realistic stereochemistry. To combat this a cross-validation factor R_{free} (Brünger, 1992b)is also used. The R_{free} is calculated from a random subset of data (the test-set) comprising 5-10% of the total data. This test-set is then omitted from refinement. Thus the deviation of F_{calc} from F_{obs} can then be measured without bias from the model. Stereochemical and conformational parameters of the model during and after structure refinement can also be used to assess the quality of the model.

Chapter 3

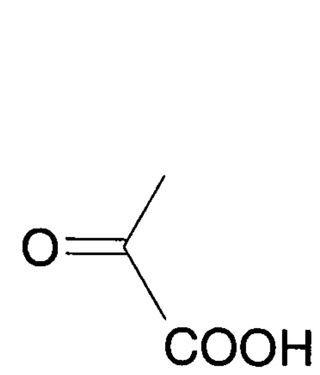
Lactobacillus delbrueckii subsp. *bulgaricus* D-2-Hydroxyisocaproate Dehydrogenase: A Broad Specificity D-2-Hydroxyacid Dehydrogenase

3.1 D-2-Hydroxyisocaproate Dehydrogenase

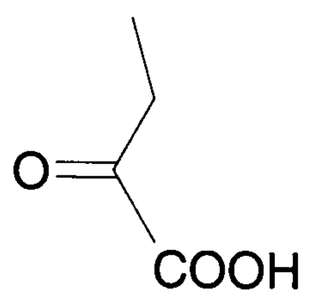
Lactobacillus delbrueckii subsp. *bulgaricus* D-2-hydroxyisocaproate (or D-2-hydroxy-4-methylpentanoate) dehydrogenase was discovered when a genomic library from *L.bulgaricus* was used to complement an *E.coli* mutant strain, grown in anaerobic conditions, deficient for both lactate dehydrogenase and pyruvate formate lyase (Bernard *et al.*, 1994).

D-2-hydroxyisocaproate dehydrogenase (D-HicDH) has been identified in several species such as *L.casei* (Hummel *et al.*, 1985), *L.curvatus* (Hummel *et al.*, 1988) and the yeast *Rhodotorula graminis* (Baker *et al.*, 1992). The enzyme has a broad substrate specificity. It catalyses a wide spectrum of D-2-hydroxyacids with branched and unbranched aliphatic side chains of medium size.

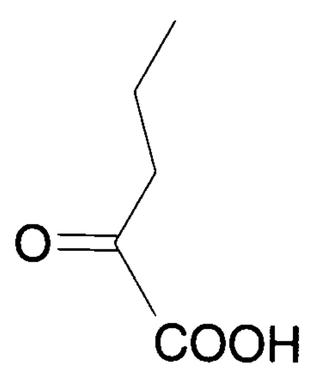
The kinetic mechanism, coenzyme binding and substrate inhibition of *L.bulgaricus* D-HicDH had previously been studied by Alvarez *et al.* (1997). They showed that the enzyme has a bi-bi ordered mechanism with NADH binding first, and that the enzyme could be inhibited by excess concentrations of pyruvate. This substrate inhibition was shown to be be directly attributed to the formation of a NAD^+-pyruvate adduct. The substrate specificity range of the enzyme has also been examined (Johnsen, 1995; Bernard *et al.*, 1994) (Table 3.1). The variety of 2-ketoacids used in these studies are shown in Figure 3.1.



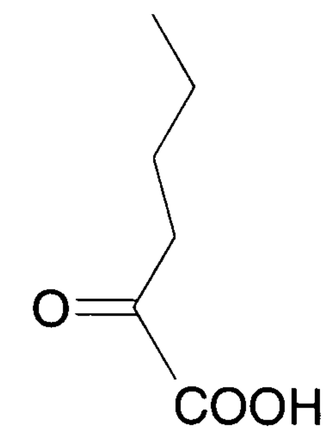
pyruvate



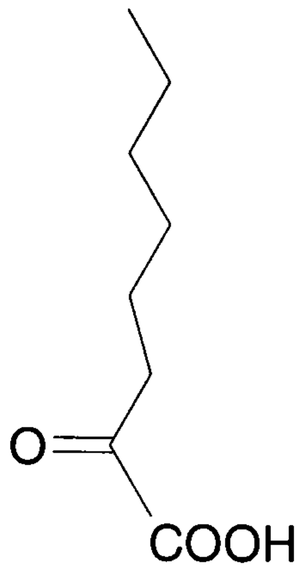
2-ketobutyrate



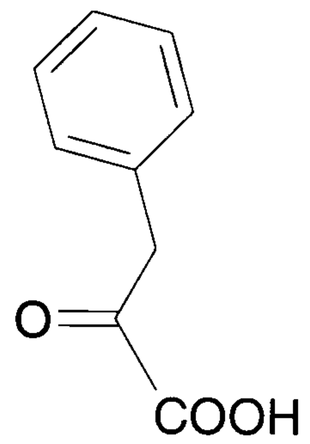
2-ketovalerate



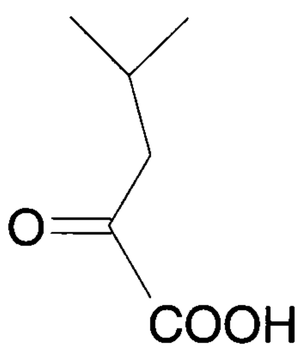
2-ketocaproate



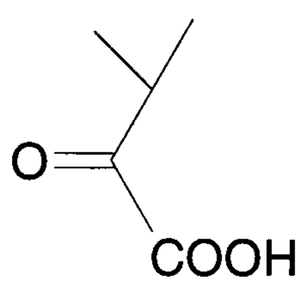
2-ketooctanoate



phenylpyruvate

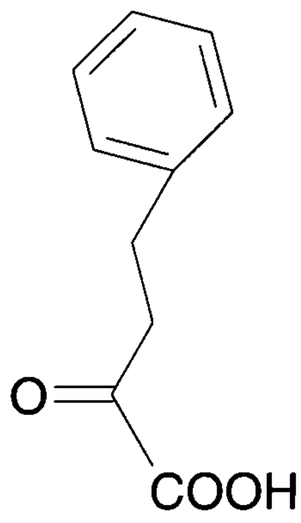


ketoisocaproate

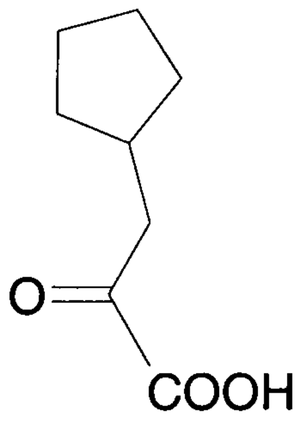


ketoisovalerate

Non-physiological substrates



benzylpyruvate



cyclopentyl pyruvate

Figure 3.1 The chemical structure of the 2-ketoacids used in D-HicDH specificity studies.

SUBSTRATE	k_{cat} (s^{-1})	K_M (μM)	K_i (mM)	k_{cat}/K_M ($s^{-1}M^{-1}$)
phenylpyruvate	22.2 ± 0.7	31.5 ± 4.0	4.7 ± 1.8	7.05×10^5
2-ketooctanoate	91.4 ± 1.4	298.6 ± 14.4	-	3.06×10^5
2-ketocaproate	14.7 ± 0.1	19.7 ± 0.6	18.2 ± 1.4	7.46×10^5
2-ketoisocaproate	21.5 ± 0.7	21.8 ± 2.9	62.4 ± 47.8	9.86×10^5
2-ketovalerate	21.1 ± 0.4	57.1 ± 4.1	33.5 ± 5.4	3.70×10^5
2-ketoisovalerate	16.0 ± 0.6	20500 ± 1600	-	7.80×10^2
2-ketobutyrate	51.6 ± 1.1	1750 ± 80	53.3 ± 5.3	2.95×10^4
pyruvate	47.5 ± 3.0	15000 ± 1900	-	3.17×10^3
benzylpyruvate	56.7 ± 1.4	36.1 ± 2.7	11.2 ± 2.8	1.57×10^6
cyclopentyl pyruvate	11.9 ± 0.2	14.1 ± 0.8	5.5 ± 1.0	8.44×10^5

Table 3.1 Steady state kinetic constants for the various substrates reduced by D-HicDH. The assay solutions contained 0.2mM NADH and were at pH 7.5. Measurements were taken at 25°C. (Data taken from Johnsen, 1995).

Most of the substrates tested weakly inhibited the enzyme, apart from pyruvate, 2-ketoisovalverate and 2-ketooctanoate. Inhibition by phenylpyruvate was stronger. The preferred substrates of *L.bulgaricus* D-HicDH are those which consist of a three-carbon 2-oxocarboxylic moiety whose C3 is substituted with a hydrophobic 'tail'. Thus, apart from C3-branched 5 carbon 2-ketoisovalerate, the substrates with strongly hydrophobic tails (namely, benzylpyruvate, cyclopentyl pyruvate, 2-ketoisocaproate, 2-ketocaproate, phenylpyruvate and 2-ketovalerate) are recognised better than the one carbon substituted 2-ketooctanoate. The unsubstitued pyruvate, along with 2-ketoisovalerate are poorly recognised substrates.

The high K_M for 2-ketoisovalerate, is revealing about the structure of the substrate binding pocket of the enzyme. It has been suggested that the pocket consists of two compartments (Johnsen, 1995). The first would be the active site of the protein where the 2-oxocarboxylic moiety would sit. The second site could be a region where hydrophobic interactions take place. The compartments

may be linked by a narrow cleft, and this could explain the poor binding of the β-branched 2-ketoisovalerate.

A sequence comparison of several D-HicDH and D-LDH enzymes (Figure 3.2) shows that accommodation of larger substrates in the broad specificity enzymes is not due to insertions in the active site loops of D-HicDH (*cf.* L-HicDH and L-LDH section 1.4.4.3).

The given name of the enzyme, 2-hydroxyisocaproate dehydrogenase, is a misnomer, as the enzyme is no more specific for 2-ketoisocaproate than for phenylpyruvate or 2-ketocaproate.

```
                          1                                                                                                  99
L.bulgaricusD-HicDH       MTKIAMYNVS PIEVPYIEDW AKKN-DVEIK TTDQALTSAT VDLAEGCSSV SLKPLGPVDE EVVYQKLSEY GVKCIGLRIV GFNTINFDWT KKYNLLVTNV
L.caseiD-HicDH             MKIIAYGAR VDEIQYFKQW AKDTGNT-LE YHTEFLDENT VEWAKGFDGI NSLQTTPYAA GVF-EKMHAY GIKFLTIRNV GTDNIDMTAM KQYGIRLSNV
L.bulgaricusD-LDH          TKIFAYAIR EDEKPFLKEW EDAHKDVEVE YTDKLLTPET AALAKGADGV VVYQQLDYTA ETL-QALADN GITKMSLRNV GVDNIDMAKA KELGFQITNV
L.helveticusD-LDH          TKVFAYAIR KDEEPFLNEW KEAHKDIDVD YTDKLLTPET AKLAKGADGV VVYQQLDYTA DTL-QALADA GVTKMSLRNV GVDNIDMDKA KELGFQITNV
L.johnsoniiD-LDH          MTKIFAYAIR KDEEPFLNEW KDAHKDIEVE YTDKLLAPET AKLAKGADGV VVYQQLDYTP ETL-QALADA GVTKMSLRNV GVDNIDMDKA KELGFEITNV
L.pentosusD-LDH            MKIIAYAVR DDERPFFDTW MKENPDVEVK LVPELLTEDN VDLAKGFDGA DVYQQKDYTA EVL-NKLADE GVKNISLRNV GVDNLDVPTV KARGLNISNV
 Consensus                 .tKI.aY..r .dE.p%...W ak...#ve.. ytd..Lt..t v.lAkG.dgv ...q..py.a ev..#k$..y G!k...lRnV G.#ni#m... K.yg...tNV

                          100                                                                                                198
L.bulgaricusD-HicDH       PVYSPRAIAE MTVTQAMYLL RKIGEFRYRM DHDHDFTWPS NLISNEIYNL TVGLIGVGHI GSGLAEIFSA MGAKVIAYDV AYNPEFEPFL TYTD-FDTVL
L.caseiD-HicDH            PAYSPAAIAE FALTDTLYLL RNMGKVQAQL QA-GDYEKAG TFIGKELGQQ TVGVMGTGHI GQVAIKLFKG FGAKVIAYDP YPMKGDHPDF DYV-SLEDLF
L.bulgaricusD-LDH         PVYSPNAIAE HAAIQAARIL RQAKAMDEKV AR-HDLRWAP T-IGREVRDQ VVGVVGTGHI GQVFMQIMEG FGAKVIAYDI FRNPELEKKG YYVDSLDDLY
L.helveticusD-LDH         PVYSPNAIAE HAAIQAARVL RQDKRMDEKM AK-RDLRWAP T-IGREVRDQ VVGVVGTGHI GQVFMRIMEG FGAKVIAYDI FKNPELEKKG YYVDSLDDLY
L.johnsoniiD-LDH          PVYSPDAIAE HAAIQAARVL RQDKRMDEKM AK-RDLRWAP T-IGREVRDQ VVGVVGTGHI GQVFMKIMEG FGAKVIAYDI FKNPELEKKG YYVDSLDDLY
L.pentosusD-LDH           PAYSPNAIAE LSVTQLMQLL RQTPMFNKKL AK-QDFRWAP D-IAKELNTM TVGVIGTGRI GRAAIDIFKG FGAKVIGYDV YRNAELEKEG MYVDTLDELY
 Consensus                PvYSP.AIAE .a.t#a.ylL R..g..... ....D..wa. t.Ig.E...q tVGv.GtGhI Gqv...if.g fGAKVIaYD. ..npe.ep.. .Yvdsl#dl.

                          199                                                                                                298
L.bulgaricusD-HicDH       KEADIVSLHT PLFPSTENMI GEKQLKEMKK SAYLINCARG ELVDTGALIK ALQDGEIAGA GLDTLAGESS YFGHTGLTDS EIPEDYKTLA KMPNVVITPH
L.caseiD-HicDH            KQSDVIDLHV PGIEQNTHII NEAAFNLMKP GAIVINTARP NLIDTQAMLS NLKSGKLAGV GIDTYEYETE DLLNLAKHGS FKDPLWDELL GMPNVVLSPH
L.bulgaricusD-LDH         KQADVISLHV PDVPANVHMI NDKSIAKMKQ DVVIVNVSRG PLVDTDAVIR GLDSGKVFGY AMDVYEGEVG VFNEDREGKE FPDARLADLI ARPNVLVTPH
L.helveticusD-LDH         KQADVISLHV PDVPANVHMI NDKSIAEMKD GVVIVNCSRG RLVDTDAVIR GLDSGKIFGF VMDTYEDEVG VFNKDWEGKE FPDKRLADLI DRPNVLVTPH
L.johnsoniiD-LDH          KQADVISLHV PDVPANVHMI NDESIAKMKD GVVIVNCSRG PLVDTDAVIR GLDSGKIFGF VMDTYEGEVG VFNEDWEGKE FPDARLADLI DRPNVLVTPH
L.pentosusD-LDH           AQADVITLHV PALKDNYHML NADAFSKMKD GAYILNFARG TLIDSEDLIK ALDSGKVAGA ALVTYEYETK IFNKDLEGQT IDDKVFMNLF NRDNVLITPH
 Consensus                k#aD!!sLHv P..p.n.hmi ne.....MK. ga..iN.aRg .L!Dt.a.i. .L.sGk.aG. g.dtye.E.. .f.......s f.d.....L. .mpNVv.tPH

                          299                                  333
L.bulgaricusD-HicDH       SAFYTETSIR NMVQICLTDQ LTIAKGRRPR SIVNL
L.caseiD-HicDH            IAYYTETAVH NMVYFSLQHL VDFLTKGETS TEVTGPAK
L.bulgaricusD-LDH         TAFYTTHAVR NMVVKAFDNN LELVEGKEAE TPVKVG
L.helveticusD-LDH         TAFYTTHAVR NMVVKAFNNN LKLINGEKPD SPVALNKNKF
L.johnsoniiD-LDH          TAFYTTHAVR NMVTKAFDNN LKMINGEKPD SPVALDKNKF
L.pentosusD-LDH           TAFYTETAVH NMVHVSMNSN KQFIETGKAD TQVKFD
 Consensus                .A%YTeta!r NMV...l... l....g.... t.V.......
```

Figure 3.2. Sequence alignment of NAD^+-dependent D-HicDHs and D-LDHs. Residues in red are highly conserved (> 90%); residues in blue are not highly conserved (> 50%). If the remaining residues are conservative substitutions the consensus residue is in uppercase. Produced using *MultAlin*.

3.2 The H205Q Mutation

In order to assign functions to specific amino acids in *L.bulgaricus* D-HicDH mutants of conserved and semi-conserved residues (Figure 1.14) were constructed and characterised (Bernard *et al.*, 1996). Enzyme inactivation with diethylpyrocarbonate indicated that a single histidine residue was involved in catalysis and it was demostrated that H296 is responsible for proton exchange in the reaction. The acid residues D259 or E264 were thought to be involved in the stabilisation of His296 in the protonated state. However mutation of both residues caused a similar and large reduction in k_{cat}/K_M. This indicated that D259 and/or E264 could partner H296. Mutation of the R235 residue lead to an increase in K_M and a reduction in k_{cat} for 2-ketoisocaproate. This residue is probably involved binding of the carboxylate group of the substrate.

A partially conserved histidine residue at 205 (conserved in 10/13 members of the D-2-hydroxyacid dehydrogenase family) was then mutated to a glutamine. The H205Q mutant showed, a 7-fold increase for K_M for NADH; a 2-10 fold pH dependent increase in k_{cat} for 2-ketoisocaproate; a 2-30 fold increase in K_M for 2-ketoisocaproate; and a 7-20 fold reduction in substrate inhibition. In order to confirm the observations that the kinetic constants for the H205Q mutant had improved Johnsen (1995) performed further steady state kinetics experiments using 3 substrates (Table 3.2 and Figure 3.3).

	2-ketoisocaproate		cyclopentyl pyruvate		benzylpyruvate
	WT	H205Q	WT	H205Q	H205Q
k_{cat} (s^{-1})	41.1 ± 0.7	138.3 ± 3.0	29.7 ± 1.04	129.0 ± 2.1	284.0 ± 6.5
K_M (μM)	19.2 ± 1.0	723.7 ± 41.8	9.7 ± 1.3	332.0 ± 14.0	603.8 ± 40.2
K_i (mM)	3.55 ± 0.29	135.8 ± 32.7	3.5 ± 0.7	54.7 ± 9.3	-
k_{cat}/ K_M ($s^{-1}M^{-1}$)	2.4×10^6	1.92×10^5	3.06×10^6	3.89×10^5	4.70×10^5

Table 3.2 Steady state kinetic constants for the various substrates reduced by WT and H205Q D-HicDH. The assay solutions contained 0.2mM NADH and were at pH 6.5. Measurements were taken at 25°C. (Data taken from Johnsen, 1995).

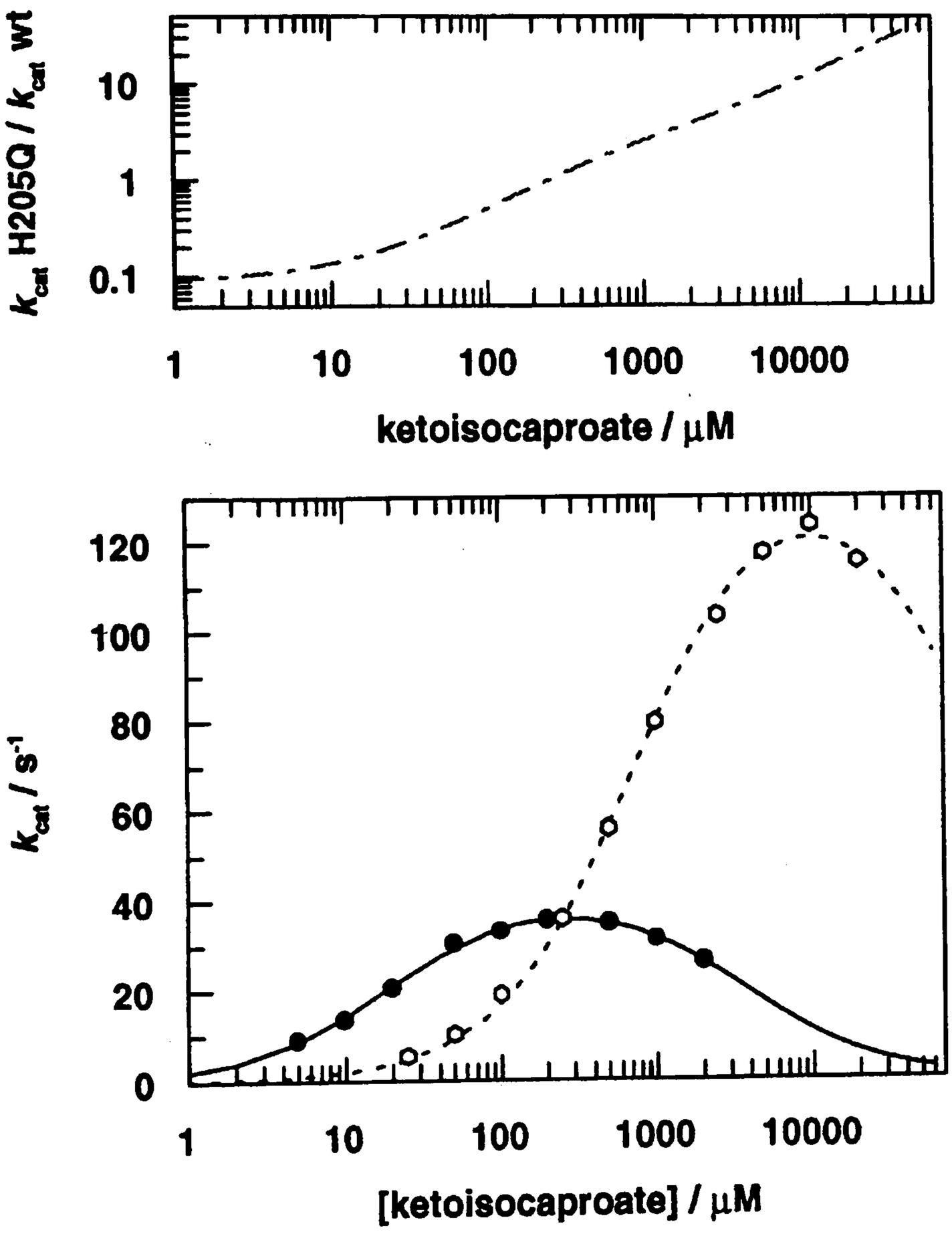


Figure 3.3 Observed and calculated k_{cat} for WT and H205Q mutant D-HicDH. Filled circles are the data points for wild type D-HicDH. Open circles are data points for H205Q mutant D-HicDH. (Similar graphs were obtained for other 2-ketoacid substrates Johnsen 1995).

The graphs show that the H205Q mutant enzyme has a vastly superior catalytic rate at high substrate concentrations. The table shows that the mutant enzyme also has higher K_M values and lower K_i values compared with wild type D-HicDH. These modifications in the kinetic properties make the H205Q mutant enzyme very attractive for use in the industrial production of chirally pure compounds. For example, the enhanced mutant enzyme may be used in the production of chirally pure L-phenylalanine, a synthon in the production of the dipeptide sweetner Aspartame (Figure 3.4).

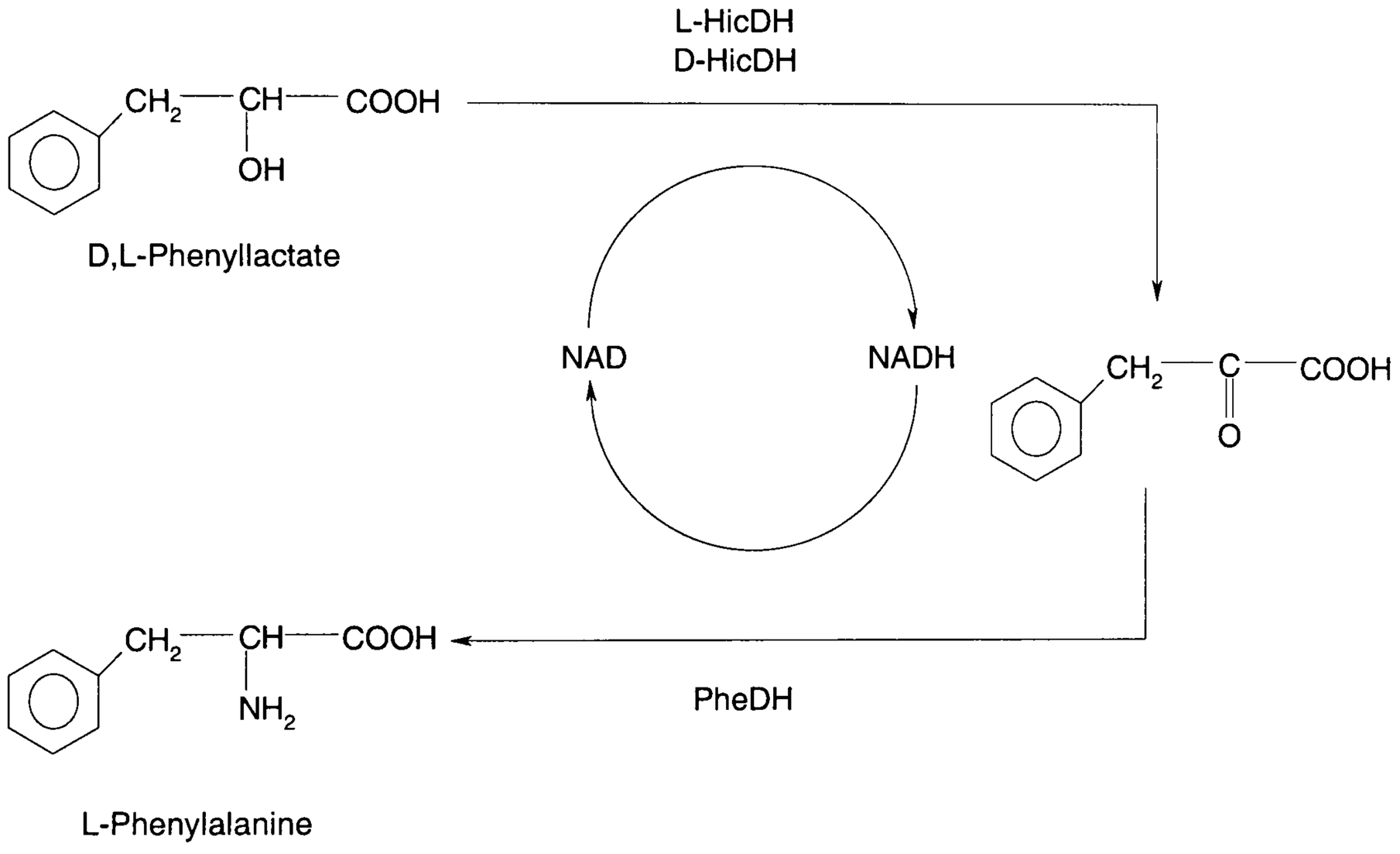


Figure 3.4 The prepararion of phenylalanine. Oxidation of D,L-phenyllactate with D- and L-HicDH and simultaneous reductive amination of the formed phenylpyruvate with phenylalanine dehydrogenase (PheDH).

Comparsion of the *L.bulgaricus* D-HicDH sequence with the crystal structures of *L.casei* D-HicDH (Dengler *et al.*, 1997) and *L.helveticus* D-LDH (Dunn and Holbrook, 1991) suggest that the H205 residue is on the βD strand of the Rossmann fold motif in the coenzyme binding site. The structures show that H205 is within 5Å of the pyrophosphate backbone of NADH. Therefore it is presumed that the reduced binding of NAD^+ and NADH to the H205Q mutant

(Bernard et al., 1996) may be the result of weakened binding between the uncharged glutamine and pyrophosphate. This increased NAD^+-off rate would therefore decrease the probability of the formation of an abortive adduct and so improve the catalytic efficiency of the mutant enzyme.

Interestingly however, the His 205 residue has recently *not* been implicated in coenzyme binding in L.bulgaricus D-LDH (Kochhar, 2000). The K_M for NADH remained essentially unchanged when His 205 was replaced with lysine.

3.3 Aims

Molecular models of the wild type and H205Q mutant *L.bugaricus* D-HicDH enzymes were required in order to determine the structural basis for substrate specificity and stereospecificity, and also to understand the role played by the H205Q mutation in enhancing the catalytic efficiency of the enzyme.

Chapter 4

Purification, Crystallisation and Diffraction Data Collection of the Wild Type & H205Q mutant D-2-Hydroxyisocaproate Dehydrogenases

L.bulgaricus D-HicDH has a low pI of 4.8 and is stable overnight at pH 4.5 (Johnsen, 1995). Using these physiochemical properties a single purification protocol had previously been devised for the wild type and mutant enzymes (Bernard *et al.*, 1994).

4.1 Purification of D-HicDH

As the pGIN113/6 plasmid carries an ampicillin resistance gene, the Luria-Bertani (LB) agar and Luria (L) Broth (Sambrook *et al.*, 1989) used to grow the *E.coli* cells all contained 1 μg/ml of ampicillin to select for the cell strain expressing the D-HicDH (wild type or H205Q) enzyme.

Using aseptic techniques, a glycerol stock of the wild type clone AR58p[GIN113/6] (or the clone harbouring the H205Q mutant) was streaked out onto a 1.6% LB agar plate and grown overnight at 37°C. A single bacterial colony was toothpicked from the plate and used to inoculate 5ml L Broth, this was again grown at 37°C in a shaking incubator (220 revolutions/min) for 8 hours. 500μl of the culture was transferred to a sterile 250ml flask containing 50ml L Broth, which was then grown for a further 4 hours at 37°C.

The 50ml L Broth was used to inoculate a sterile 2 litre flask containing 500ml L Broth, this was then grown at 37°C. When the culture had reached mid-log phase of growth (usually A_{600nm} 0.4-0.6) protein expression was induced by derepressing the lambda repressor (Turner *et al.*, 1997); this was achieved by

derepressing the lambda repressor (Turner *et al.*, 1997); this was achieved by heat shocking the culture at 42°C for 3 hours. After induction, the bacterial culture was grown at 37°C overnight.

The cells were then harvested by centrifugation at 4,000rpm for 20 minutes at 4°C. Following this, the cells were resuspended in a minimal volume of 50mM TEA pH 6.5. The bacterial suspension was then sonicated on ice, using 5 repeats of 10 seconds medium-power sonication bursts. The subsequent cell debris was pelleted by centrifugation at 4°C for 25 minutes at 20,000rpm. The supernatant was then dialysed against 50mM sodium citrate pH 4.5 overnight.

The dialysate was spun at 20,000rpm for 20 minutes at 4°C in order to remove any insoluble material. The supernatant was loaded onto a column packed with 250ml S- Sepharose resin (Sigma) which had been pre-equilibrated with 2 column volumes of 50mM sodium citrate pH 4.5. The unbound protein fraction was eluted by washing the cationic exchange column with a further 2 column volumes of 50mM sodium citrate pH 4.5 (Figure 4.1).

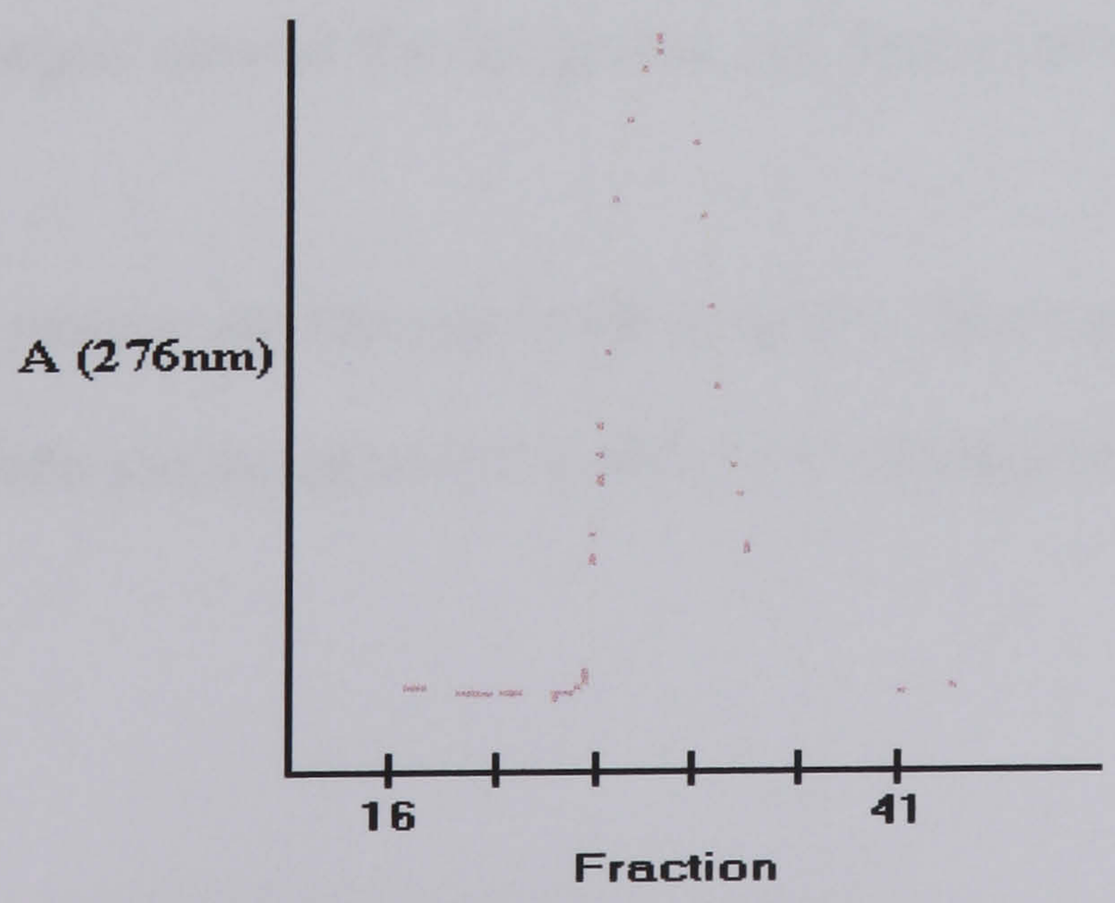


Figure 4.1 Plot of unbound fraction eluting from the S- Sepharose column.
The column was run 'under gravity' and the elution profile was followed by measuring A_{276nm} using a Pharmacia UVICORD (8300 Uvicord II) and a chart recorder.

The unbound fraction was then dialysed against 50mM triethanolamine (TEA) pH 6.5 overnight.

The protein sample was again spun at 20,000rpm for 20 minutes at 4°C to get rid of any sediment and the supernatant loaded onto a column packed with 150ml Q- Sepharose resin (Sigma) which had been equilibrated with 2 column volumes of 50mM TEA pH 6.5. The bound protein fraction was then eluted with a 600ml 0 - 0.5M NaCl gradient (Figure 4.2).

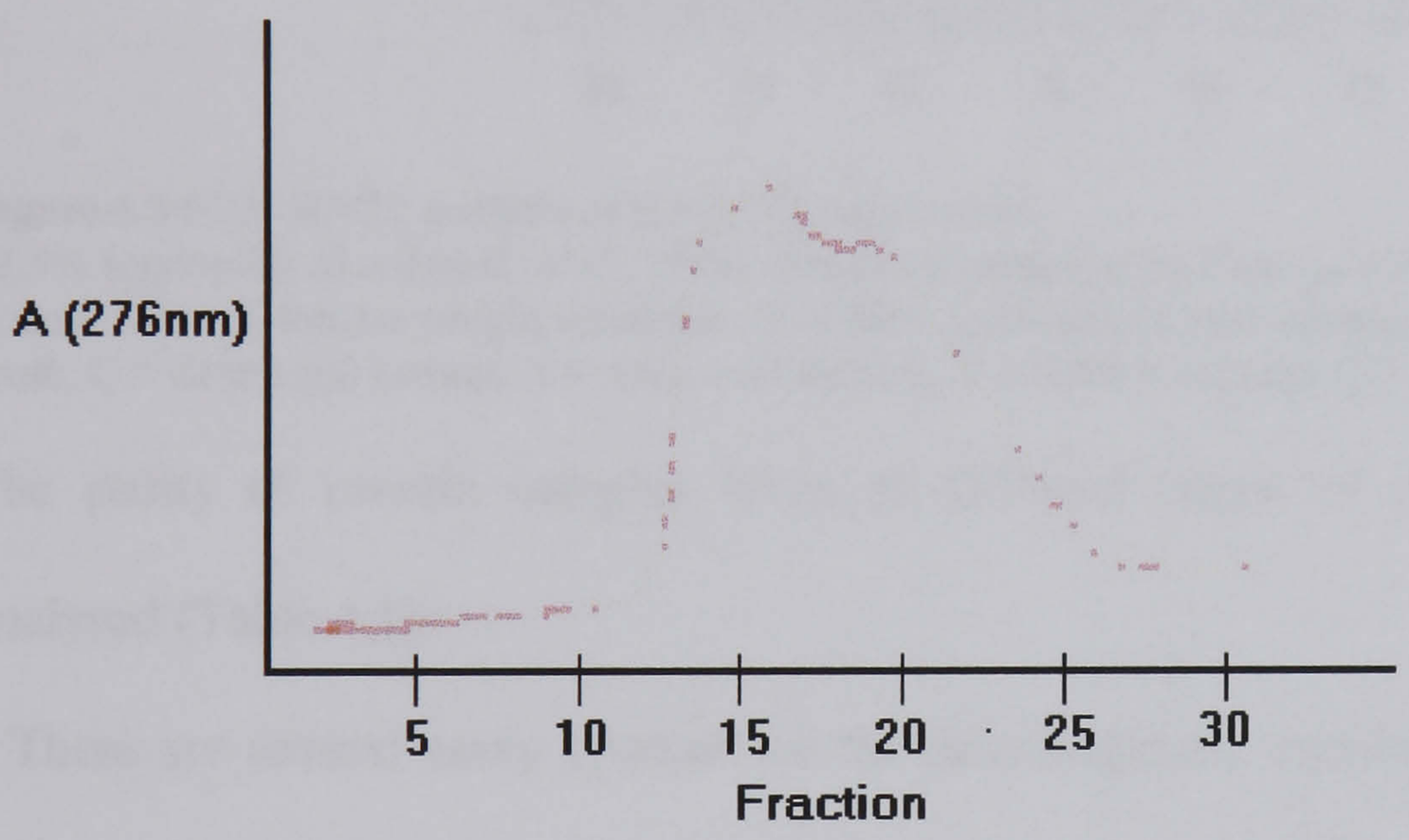


Figure 4.2 Plot of bound fraction eluting from the Q- Sepharose column.
The column was run at a flow rate of 0.6ml ml/min overnight. The elution profile was followed by measuring A_{276nm} . The D-HicDH protein eluted at 0.25-0.3M NaCl.

SDS-PAGE analysis showed that the protein had been purified to homogeneity (Figure 4.3).

The purified protein was then sterilised using a 0.22µm syringe driven filter.

Finally, the protein was stored at 4°C in 40% (w/v) ammonium sulphate.

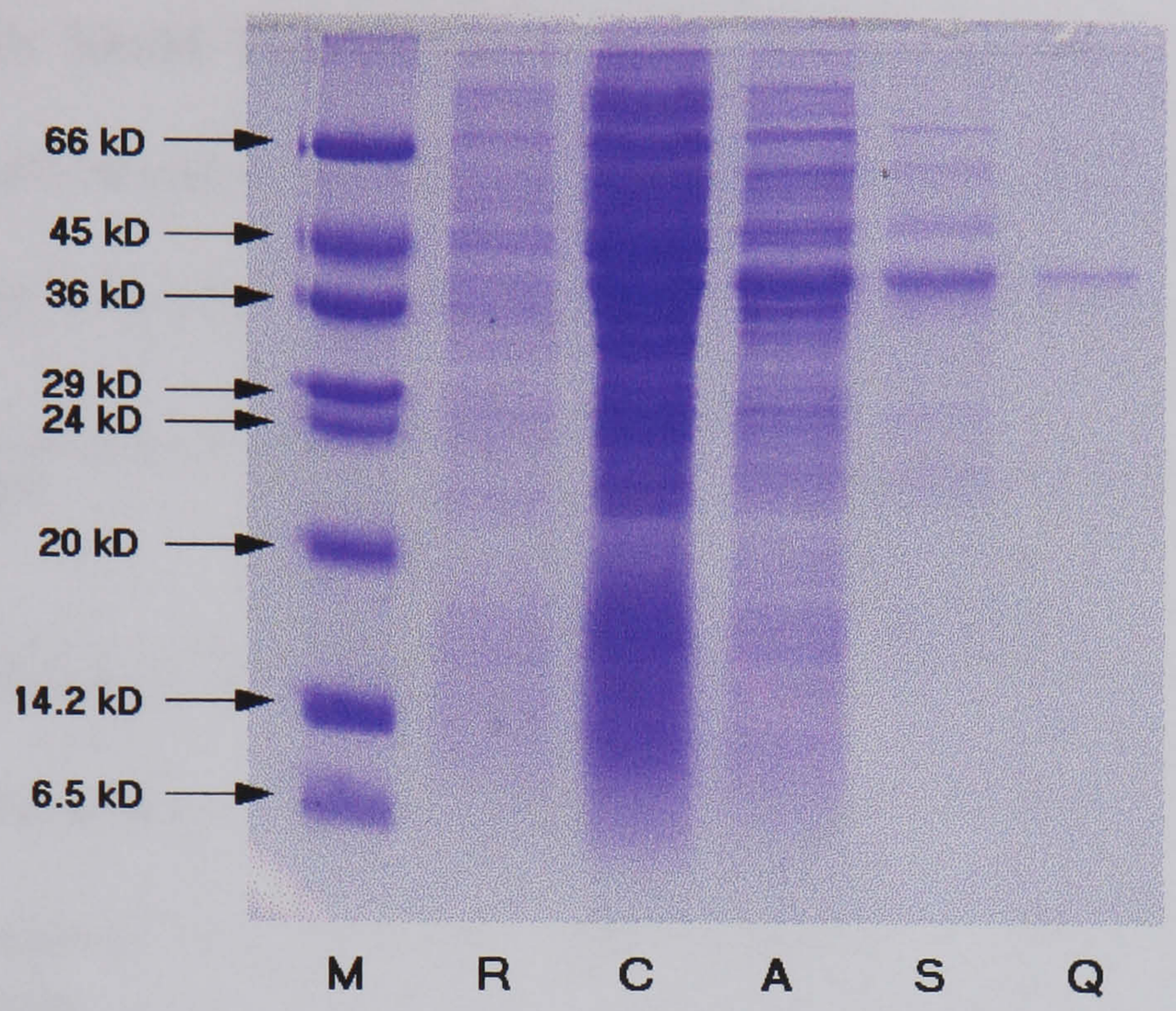


Figure 4.3 SDS- PAGE analysis of D-HicDH purification.
12.5% acrylamide (Sambrook *et al.*, 1989). Bands visualised using Comassie blue stain.
(Lanes: M = Molecular weight standards, R = Rich crude cell extract sample taken from 50ml L broth, C = Crude cell extract, A = After acid dialysis, S = After S-column, Q = After Q-column).

The purity of protein samples taken at different stages of purification was analysed (Table 4.1).

There are several assay systems for the dehydrogenase enzymes (Holbrook *et al.*, 1975). The assay used involved spectrophotometrically following the change in absorbance at 340nm (Kornberg, 1955): NADH absorbs light at 340nm, however. this is removed by hydride transfer during the catalytic forward reaction (ie. the oxidised form, NAD^+, does not absorb at 340nm). Therefore it is easy to follow any enzyme reaction consuming or generating NADH.

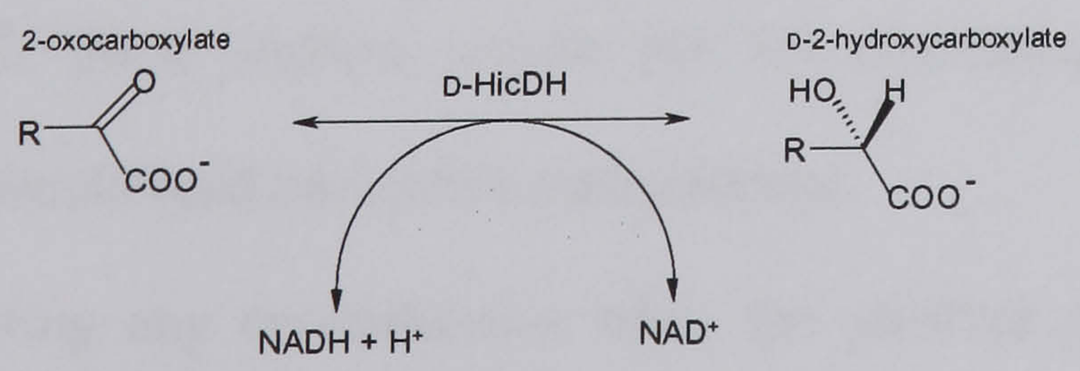


Figure 4.4 The broad specificity reaction catalysed by D-HicDH.
The reaction equilibrium is in favour of the reduction of 2-keto acid to D-2-hydroxyacid.

The assay reaction mixture consisted of 0.2mM NADH, 1mM α-ketoisocaproate buffered with 50mM TEA pH 8.0 to which specific volumes of the enzyme solution would be added (the total reaction volume being 1ml). Steady state rate measurements were taken from the initial linear portion of the reduction graph.

Stage	Volume (ml)	Total protein (mg) [a]	Total activity (U) [b]	Specific activity (U/mg)	Yield (%)	Purification factor -fold
Crude cell extract	84	614	3506	5.7	100	1.0
After acid dialysis	79	297	3484	11.7	99.4	2.1
After cationic exchange (S - column)	111	182	2981	16.4	85.0	2.9
After anionic exchange (Q - column)	153	114	2589	22.7	73.8	4.0

Table 4.1 Typical D-HicDH purification statistics from a 2 litre L Broth preparation.
[a] Protein concentration determined using the Bradford method (Bradford, 1976). This method was used instead of measuring A_{280nm} , as nucleic acids also absorb strongly at 280nm.
[b] A unit (U) is the amount of enzyme required to catalyse the oxidation of 1μmol of NADH / min at 25°C.

4.2 Protein Concentration

Firstly, any salts (e.g. ammonium sulphate, NaCl, etc.) and bound nucleotides (e.g. NAD^+) present in the sample were removed as they would effect the crystallisation of the enzyme: This was achieved by dialysing the protein against large volumes of 5mM sodium acetate pH 5.0 containing 1 g/l activated charcoal, which would bind nucleotide contaminants.

Before conducting any crystallisation trials the purified protein had to be concentrated: An Amicon Centricon centrifugal filter, with a molecular weight cutoff (MWCO) of 10kD, was used for this purpose. If further concentration was

necessary, an Amicon Microcon (for sample volumes less than 0.5ml) with a MWCO 10kD filter was utilised.

Protein concentration and the presence of NAD^+ were measured by recording the absorbance spectrum of the sample.

The extinction coefficient at 280nm in a 1cm pathlength cuvette for a 1mg/ml solution of D-HicDH was 0.96 (Johnsen, 1995). Protein molecules absorb almost twice as strongly at 280nm than at 260nm (Roth, 1968). However, NAD^+ has a very strong ultraviolet absorption peak at 260nm (although its reduced form, NADH, absorbs weakly at 260nm) (Holbrook *et al.*, 1975). Therefore a A_{280nm}/A_{260nm} ratio greater than 1.9 for the sample was taken to imply the absence of NAD^+ in the protein solution.

4.3 Crystallisation and Diffraction Data Collection of the Wild Type & H205Q mutant D-2-Hydroxyisocaproate Dehydrogenases

4.3.1 Crystallisation Method

Unless otherwise stated, all crystallisation trials were performed using the hanging drop variant of the vapour diffusion method, with a drop consisting of 3μl protein solution + 3μl well solution over a 1ml well, and the crystal trays were stored at 17°C.

4.3.2 Data Processing and Reduction

The X-ray data collected were processed and reduced using the HKL suite of programs (Otwinowski and Minor, 1996): The diffraction images were indexed using the *DENZO* program. The aim of 'auto-indexing' is to determine the space group, calculate the approximate unit cell dimensions and find the orientation matrix of the unit cell. This operation requires only about 100 reflections from a single diffraction image and the X-ray beam position and crystal-to-detector position must be known; these spot positions and parameters are then refined (along with other parameters, such as background corrections). The intensities of the spots can then be integrated.

The indexed data were then scaled and merged using *SCALEPACK*. The scaling operation put the intensities from different images on a common scale to account for variations in incident beam intensity, wavelength, *etc.* (see 5.2.4.1). The average mosaicity of the crystal is also calculated .The scaled data are then merged, i.e. multiple observations are reduced to a mean intensity and standard deviation.

The *scalepack2mtz* program was then used to convert the reduced data to mtz format; the format used by the CCP4 suite of progams (Collaborative Computational Project Number 4, 1994). The structure factor amplitudes were then calculated as the positive square root of the intensities using theCCP4 program *TRUNCATE*.

4.3.3 Crystallising the wild type & H205Q apoenzyme

Three crystallisation conditions were found for the H205Q mutant apoenzyme using Hampton Screen (I). These were then further optimised.

Apoenzyme crystallisation condition (1): 50mg/ml protein; 28% polyethylene glycol (PEG) 400, 0.1M Na Hepes pH 7.5, 0.2M $CaCl_2$.

[Protein solution] (mg/ml)	[PEG 400] Range (%)	pH Range	[$CaCl_2$] (M)	Results (After 5 days)
50	24 - 30	7.3 - 7.7	0.2	Small rod-shaped crystals (~ 50μm in length) grew around precipitated protein
22	2 - 30	7.3 - 7.7	0.2	Light precipitate
60	24 - 30	7.3 - 7.7	0.2	Heavy precipitate
22	16 - 20	7.3 - 7.7	0.2	Clear drop
11	24 - 30	7.3 - 7.7	0.2	Precipitate
6	24 - 30	7.3 - 7.7	0.2	Micro-clusters of crystals^{+}
45	24 - 30	7.5 - 7.7	0.2	Very small rod-shaped crystals grew around precipitate
6	10 - 16	7.3 - 7.7	0.2	Micro-clusters of crystals^{+}
17	4 - 10	7.3 - 7.7	0.2	Micro-clusters of crystals^{+}
-	4 - 10	-	0.05 - 0.5	Micro-clusters of crystals$^{\#}$

Table 4.2 A selection of trials used to optimise apoenzyme crystallisation condition (1).
$^{+}$ The Hampton Screen protein dye 'Izit' indicated that the micro-clusters were not protein.
$^{\#}$ Micro-clusters still formed in the complete absence of protein, and became more abundant at higher salt concentrations: This suggested that the clusters were $CaCl_2$ crystals.

It was found that protein crystals only formed in a very narrow range when using these conditions.

Figure 4.5 Protein crystals grown using apoenzyme crystallisation condition (1).

Apoenzyme Crystal form (1)

One crystal initially diffracted to ~3Å. However, the crystal started to rapidly deteriorate and so only a few images were collected.

Apoenzyme crystallisation condition (2): 50mg/ml protein; 20% PEG 4000, 10% propan-2-ol, 0.1M Na Hepes pH 7.5.

[Protein solution] (mg/ml)	[PEG 4000] Range (%)	[Propan-2-ol] Range (%)	pH Range	Results
49	18 - 22	10	7.3 - 7.7	Crystals grew in drops at higher pH with lower [PEG] after 2 weeks
47	14 -26	10	7.7 - 8.6	Crystals were found at mid [PEG] at pH 8.1 - 8.4
25	14 -26	10	7.7 - 8.6	All drops remained clear
47	17	2.5 - 15	7.7 - 8.6	Tiny crystals grew in precipitate at high [propan-2-ol]. However, the drops remained clear at lower concentrations
40 - 50	15 - 20	10	8.3	Crystals formed in the higher [protein] drops

Table 4.3 A selection of trials used to optimise apoenzyme crystallisation condition (2).

Figure 4.6 Protein crystals grown using apoenzyme crystallisation condition (2).

Apoenzyme Crystal form (2)

Data collected using the in-house rotating-anode X-ray source (λ = 1.5418Å).

Cryoprotectant solution	No cryoprotectant needed
Unit cell	a=102.15Å, b=113.01Å, c=102.43Å $\alpha=\gamma=90°$, $\beta=90.03°$
Space group	$P2_1$
Total number of unique reflections	15,909
Mean Redundancy	1.6
Resolution	3.3Å
Mosaicity	1.57°
Completeness - Overall (20.0-3.3Å) Outer shell (3.65-3.3Å)	82.6% 79.3%
I / σ (I) - Overall Outer shell	3.2 5.9
R_{merge} - Overall Outer shell	26.7% 36.9%

Table 4.4 Diffraction data collection statistics of a apoenzyme crystallisation condition (2) crystal.

Apoenzyme crystallisation condition (3): 50mg/ml protein; 2.0M ammonium dihydrogen phosphate, 0.1M Tris HCl pH 8.5.

[Protein solution] (mg/ml)	[$NH_4H_2PO_4$] Range (M)	pH Range	Results
50	0.25 - 1.5	8.3 - 8.7	Very small bipyramidal tetragonal crystals at higher [$NH_4H_2PO_4$] appeared after 4 days
26	0.25 - 1.5	8.3 - 8.7	Drops remained clear
30 - 50	1.5 - 2.5	8.5	Small crystals grew in drops at mid [protein] with ~2M $NH_4H_2PO_4$
45	1.0 - 2.0	8.5	Small bipyramidal tetragonal crystals 1.8-2M $NH_4H_2PO_4$ grew after 1 week

Table 4.5 A selection of trials used to optimise apoenzyme crystallisation condition (3).

Many attempts were made at increasing the size of the crystals, including increasing the hanging drop size and growing the crystals at 4°C, all proved to be unsuccessful.

In most cases, ammonium sulphate could replace ammonium dihydrogen phosphate as the precipitant and this still gave similar small tetragonal bipyramidal crystals.

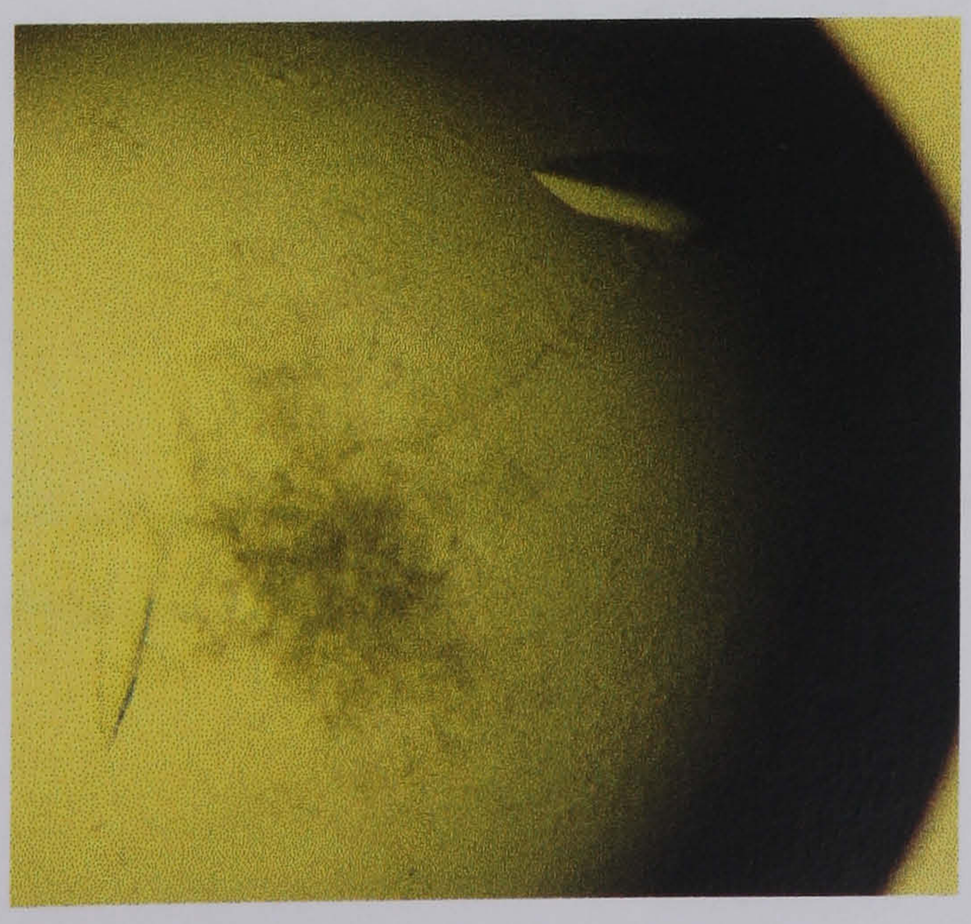


Figure 4.7 Protein crystals grown using apoenzyme crystallisation condition (3).

Virtually identical crystallisation conditions were found for the wild type apoenzyme.

Apoenzyme Crystal form (3)

The diffraction from this crystal form was poor (~4.8Å), so data were not collected.

4.3.4 Crystallising the wild type & H205Q binary complexes

The conditions used to grow binary complex (enzyme + cofactor) crystals were found to be different from apoenzyme crystallisation conditions, apart from the latter condition, using ammonium dihydrogen phosphate & 0.1M Tris HCl pH 8.5 , which only gave small irregularly shaped crystals.

Again Hampton Screen (I) was used; this gave 2 preliminary crystallisation conditions for the H205Q putative binary complex.

Binary complex crystallisation condition (1): 30mg/ml protein, 2mM NADH; 20% PEG 3350, 0.1M TEA pH 7.5.

[Protein solution] (mg/ml)	[PEG 3350] Range (%)	pH Range	Results
20 - 50	15 - 30	7.5	Tiny crystals, slightly irregular in shape, formed in the higher [protein] drops
20 - 50	25 - 40	7.5	Most drops were either clear or contained light precipitate
50 - 70	15 - 30	7.5	Some well developed crystals were found in the high [protein] drops at 26 - 28% PEG 3350
70 - 90	15 - 30	7.5	Heavy precipitate
67	24 - 30	6.8 - 7.1	Several well formed crystals in tray amongst larger irregularly shaped crystals

Table 4.6 A selection of trials used to optimise binary complex crystallisation condition (1).

The crystals were found to be unstable; after a week they would become cracked and/or lose their birefrigency. Therefore the crystals had to be frozen in liquid nitrogen in matter of days. However, whilst trying to mount the crystals, they were found to be very fragile and would sometimes crack during manipulation.

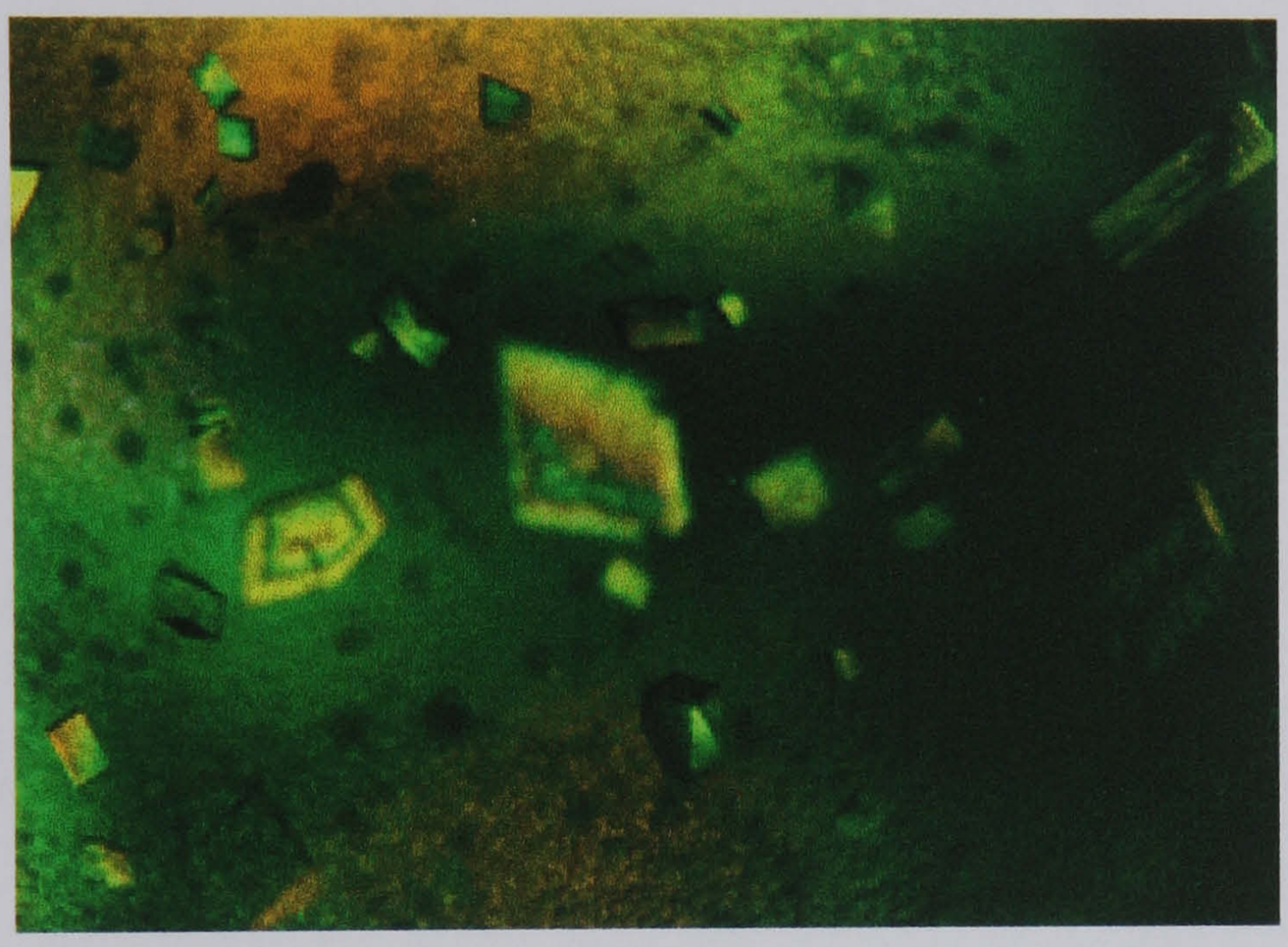


Figure 4.8 Protein crystals grown using binary complex crystallisation condition (1).

Binary complex Crystal form (1)

Data collected using the Daresbury SRS laboratory (PX7.2, λ = 1.488Å).

Cryoprotectant solution	22% PEG 3350, 0.1M TEA pH 7.1, 2mM NADH, 10% glycerol
Unit cell	a=63.32Å, b=76.31Å, c=80.64Å, α=104.23°, ß=111.61°, γ=102.72°
Space group	P1
Total number of unique reflections	15,711
Mean Redundancy	2.2
Resolution	3.1Å
Mosaicity	2.42°
Completeness - Overall (20.0-3.1Å) Outer shell (3.21-3.1Å)	79.3% 84.1%
I / σ (I) - Overall Outer shell	3.1 6.7
R_{merge} - Overall Outer shell	12.7% 26.3%

Table 4.7 Diffraction data collection statistics of a binary complex crystallisation condition (1) crystal.

Binary complex crystallisation condition (2): 15mg/ml protein, 1mM NADH; 0.4M ammonium dihydrogen phosphate.

[Protein solution] (mg/ml)	[$NH_4H_2PO_4$] Range (%)	pH Range	Drop size (µl + µl)	Results (After 5 days)
5 -20	0.2 - 2.0	-	3 + 3	Larger crystals grew at [protein] above 10mg/ml
12.5	0.2 - 2.0	-	3 + 3	Concentrations above 1.2M gave very large cubic-like crystals (~400µm in length)
25	0.2 - 2.0	-	3 + 3	Small crystals formed
10	1.2 - 2.0	4.0 - 7.5	3 + 3	Addition of a buffer caused the formation of smaller crystals.
5 - 15	1.2 - 2.0	-	6 + 6	Very small crystals grew

Table 4.8 A selection of trials used to optimise binary complex crystallisation condition (2).

Small crystals were obtained when ammonium sulphate was used as the precipitant.

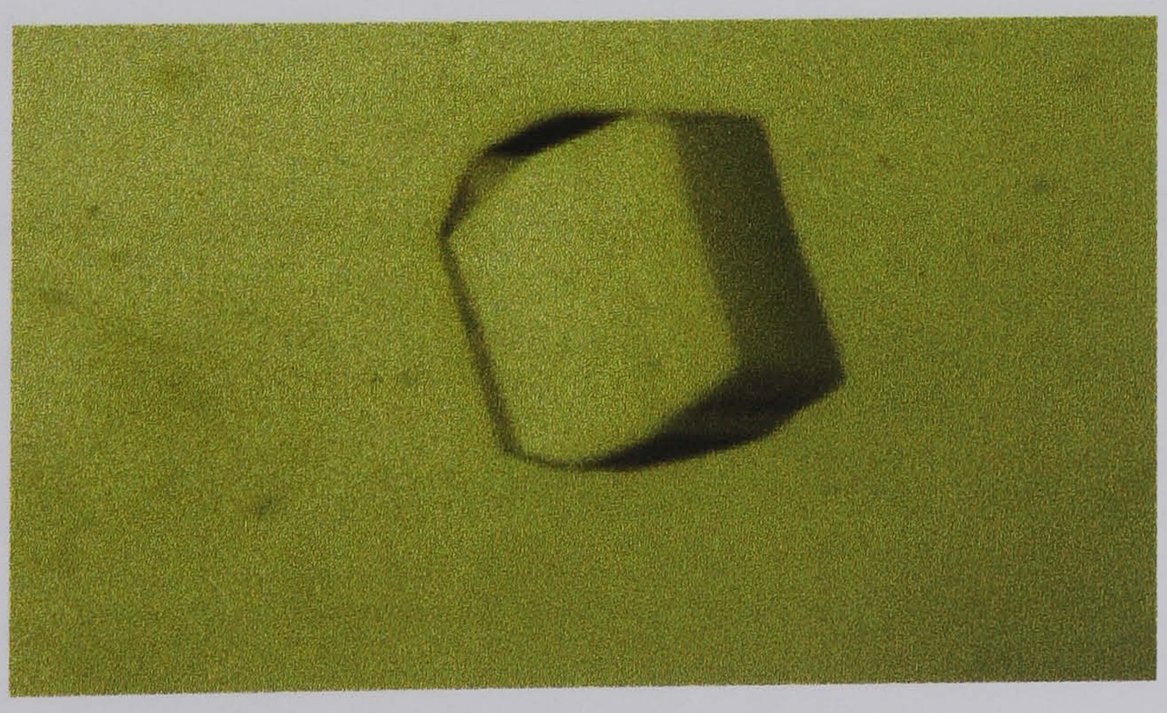


Figure 4.9 Protein crystals grown using binary complex crystallisation condition (2).

Data collected using the in-house rotating anode X-ray source.

Cryoprotectant solution	1.6M ammonium dihydrogen phosphate, 1mM NADH, 25% glycerol
Unit cell	a=b=80.06Å, c=266.95Å, α=ß=γ=90°
Space group	P422
Total number of unique reflections	7,785
Mean Redundancy	1.7
Resolution	3.6Å
Mosaicity	1.3°
Completeness - Overall (25.0-3.6Å) Outer shell (3.76-3.6Å)	84.7% 80.2%
I / σ (I) - Overall Outer shell	3.6 6.2
R_{merge} - Overall Outer shell	36.7% 44.4%

Table 4.9 Diffraction data collection statistics of a binary complex crystallisation condition (2) crystal.

One crystallisation condition was also found for the H205Q binary complex using Hampton Screen (II). Binary complex crystallisation condition (3): 30mg/ml protein, 1mM NADH; 30% PEG monomethylether (MME) 5000, 0.2M ammonium sulphate, 0.1M MES pH 6.5.

[Protein solution] (mg/ml)	[PEG MME 5000] Range (%)	pH Range	$[(NH_4)_2SO_4]$ Range (M)	Results (After ~2 weeks)
30	20 - 28	6.3 - 6.7	0.2	Rhombohedral shaped crystals grew, they decreased in size as the [PEG] increased
25	18 - 22	6.3 - 6.7	0.2	Huge crystals in drops containing 20 - 22% PEG MME 5K (Fig. 4.9a)
15 - 25	20 -22	6.3 - 6.7	0.2	Crystals decreased dramatically in size as the [protein] decreased
40	20 -22	6.3 - 6.7	0.2	Large rhombohedral shaped crystals formed in most of the drops (Fig. 4.10).However, some drops also contained oblong shaped crystals (Fig. 4.11)
50 -60	20 - 22	6.5	0.2	Large irregularly shaped crystals developed
5 - 15	24 -36	6.5	0.2	Most drops remained clear, but tiny crystals were seen others
40	20 - 22	6.5	0.05 - 0.5	Drops containing less than 0.2 M $(NH_4)_2SO_4$ remained clear. Drops which had greater than 0.2M $[(NH_4)_2SO_4]$ formed small crystals
40	20 - 22	4.5 - 5.5 using 0.1M Na acetate	0.2	The drops at pH 4.5 - 5.0 contained precipitate. The drops at pH 5.5 contained small irregularly shaped crystals and precipitate.

Table 4.10 A selection of trials used to optimise binary complex crystallisation condition (3). Varying the size of the hanging drop had negligible effect. Also, crystals did not grow when the NADH concentration was increased.

During diffraction data collection it was found that the rhombohedral and oblong shaped crystals were two distinct crystal forms (*cf.* Fig.4.10 and Fig. 4.11).

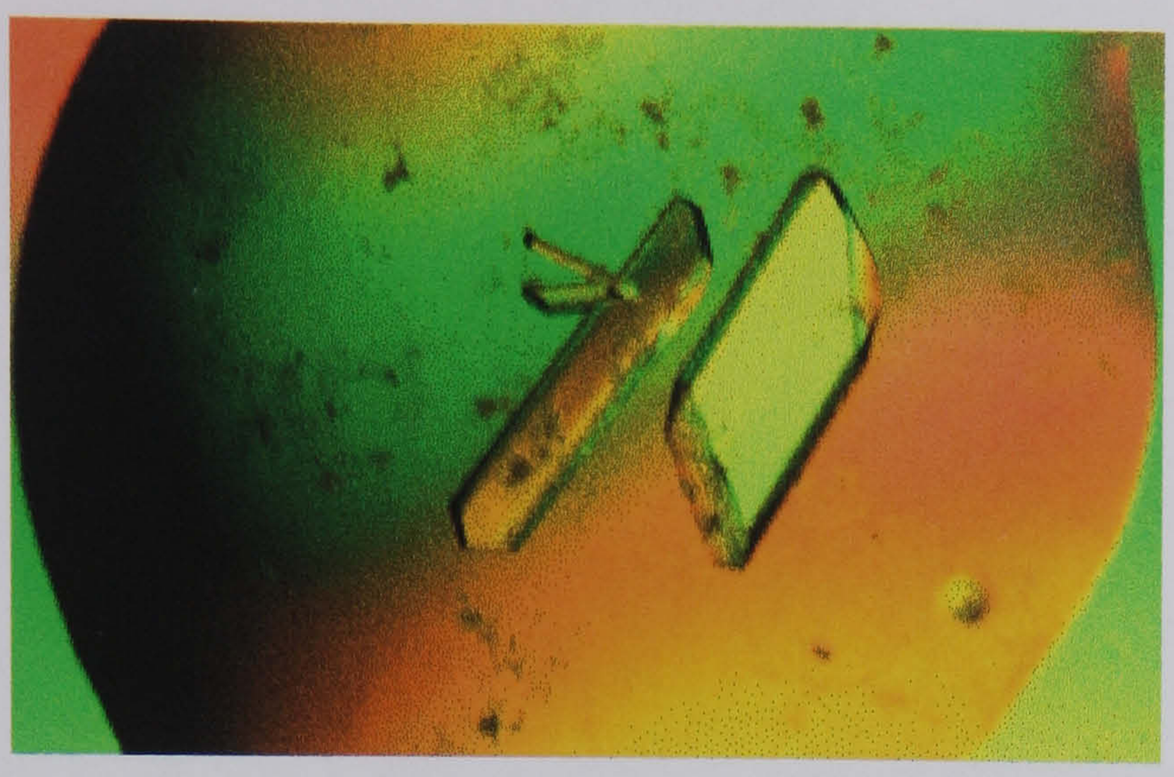


Figure 4.10 Protein crystals grown using binary complex crystallisation condition (3a).

Binary complex Crystal form (3a) $P2_1$

Data collected at the Daresbury SRS laboratory (PX9.6, $\lambda = 0.87$Å)

Cryoprotectant solution	e.g. 22% PEG MME 5K, 0.2M $(NH_4)_2SO_4$, 0.1M MES pH 6.5, 1mM NADH, 10% glycerol.
Unit cell	a=77.41Å, b=112.21Å, c=81.14Å, $\alpha=\gamma=90°$, $\beta=96.806$
Space group	$P2_1$
Total number of unique reflections	21,216
Mean Redundancy	2.09
Resolution	3.0Å
Mosaicity	0.62°
Completeness - Overall (15.00-3.00Å) Outer shell (3.24-3.00Å)	94.6% 97.1%
I / σ (I) - Overall Outer shell	3.42 11.88
R_{merge} - Overall Outer shell	5.8% 19.9%

Table 4.11 Diffraction data collection statistics of a binary complex crystallisation condition (3a) crystal.

Figure 4.11 Protein crystals grown using binary complex crystallisation condition (3b).

Binary complex Crystal form (3b) C2

Data collected at the Daresbury SRS laboratory (PX7.2)

Cryoprotectant solution	e.g. 24% PEG MME 5K, 0.2M $(NH_4)_2SO_4$, 0.1M MES pH 6.5, 1mM NADH, 10% glycerol.
Unit cell	a=161.28Å, b=80.99Å, c=125.84Å, $\alpha=\gamma=90°$, ß=118.30
Space group	C2
Total number of unique reflections	13,840
Mean Redundancy	2.4
Resolution	2.8Å
Mosaicity	0.83°
Completeness - Overall (15.0-2.8Å) Outer shell (2.94-2.8Å)	91.6% 94.3%
I / σ (I) - Overall Outer shell	3.0 7.9
R_{merge} - Overall Outer shell	8.4% 12.6%

Table 4.12 Diffraction data collection statistics of a binary complex crystallisation condition (3b) crystal.

Crystallisation trays which had been stored at 4°C contained crystals that had a slightly different external morophology and were larger than their room temperature counterparts.

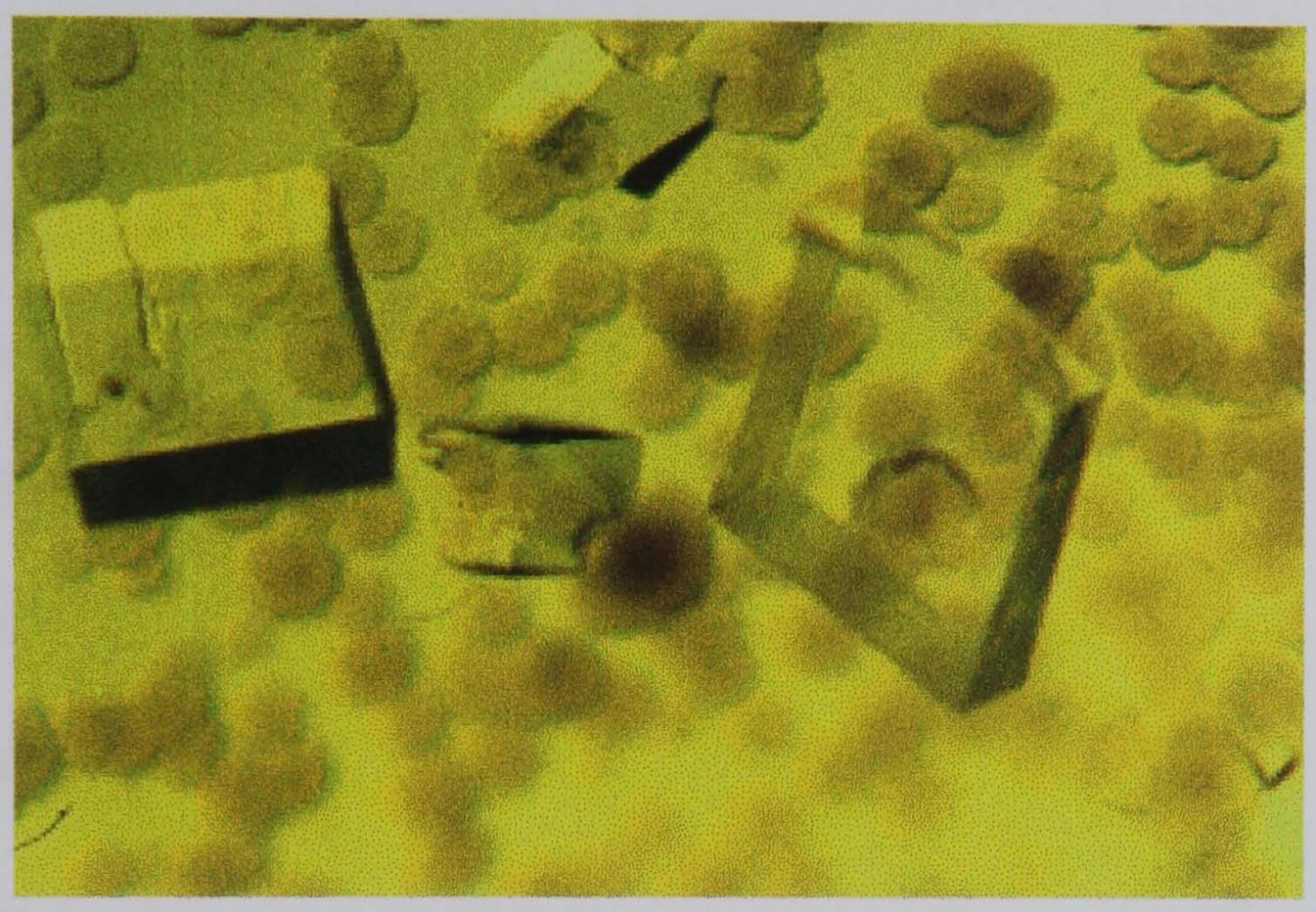


Figure 4.12 Protein crystals grown using binary complex crystallisation condition (3c).

Binary complex Crystal form (3c)

The binary crystals grown at 4°C diffracted poorly (~6Å), and so data were not collected.

The wild type putative binary complex crystallised in the same manner as the H205Q mutant using the conditions above.

4.3.5 Crystallising the wild type & H205Q ternary complexes

4.3.5.1 1,4,5,6-Tetrahydronicotinamide Adenine Dinucleotide

In order to crystallise a ternary complex (enzyme + cofactor + substrate), the catalytic 'turnover' of substrate & cofactor had to be prevented. For this purpose, 1,4,5,6- tetrahydronicotinamide adenine dinucleotide (tNADH) (Dave *et al.*, 1968) was used: tNADH, a gift from T.Dafforn (Department of Biochemistry, Bristol University), is a synthesised highly reduced inactive analogue of NADH.

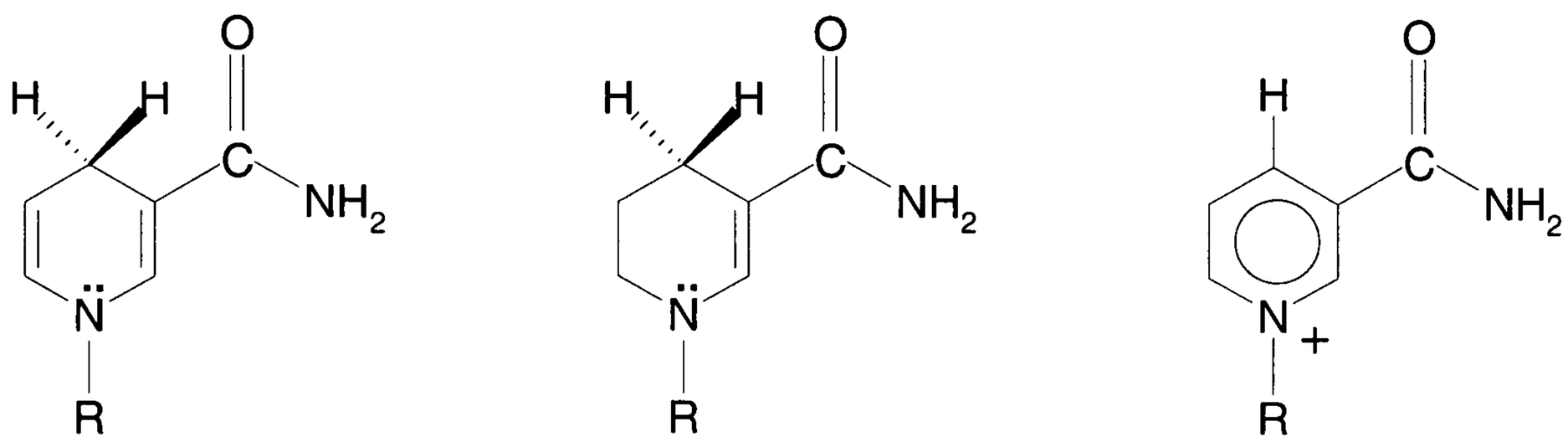


Figure 4.13 tNADH is an inactive analogue of NADH.

Previous studies on other dehydrogenases have shown that the analogue has similar binding affinities to the enzymes as NAD^+ (Thevenot *et al.*, 1975) and that tNADH binds to deydrogenases in a very similar conformation to NADH (Cedergren-Zappezauer *et al.*, 1982).

4.3.5.2 Crystallising the abortive ternary complex

Only one of the previous crystallisation conditions (binary condition(2)) yielded suitable H205Q D-HicDH putative ternary complex crystals:-

Ternary complex crystallisation condition (1): 8 - 12mg/ml protein, 1mM tNADH, 2.5mM α-ketoisocaproate in the drop solution; 1.2 - 1.4M ammonium dihydrogen phosphate, 0.1M Na acetate pH 5.0 in the well solution.

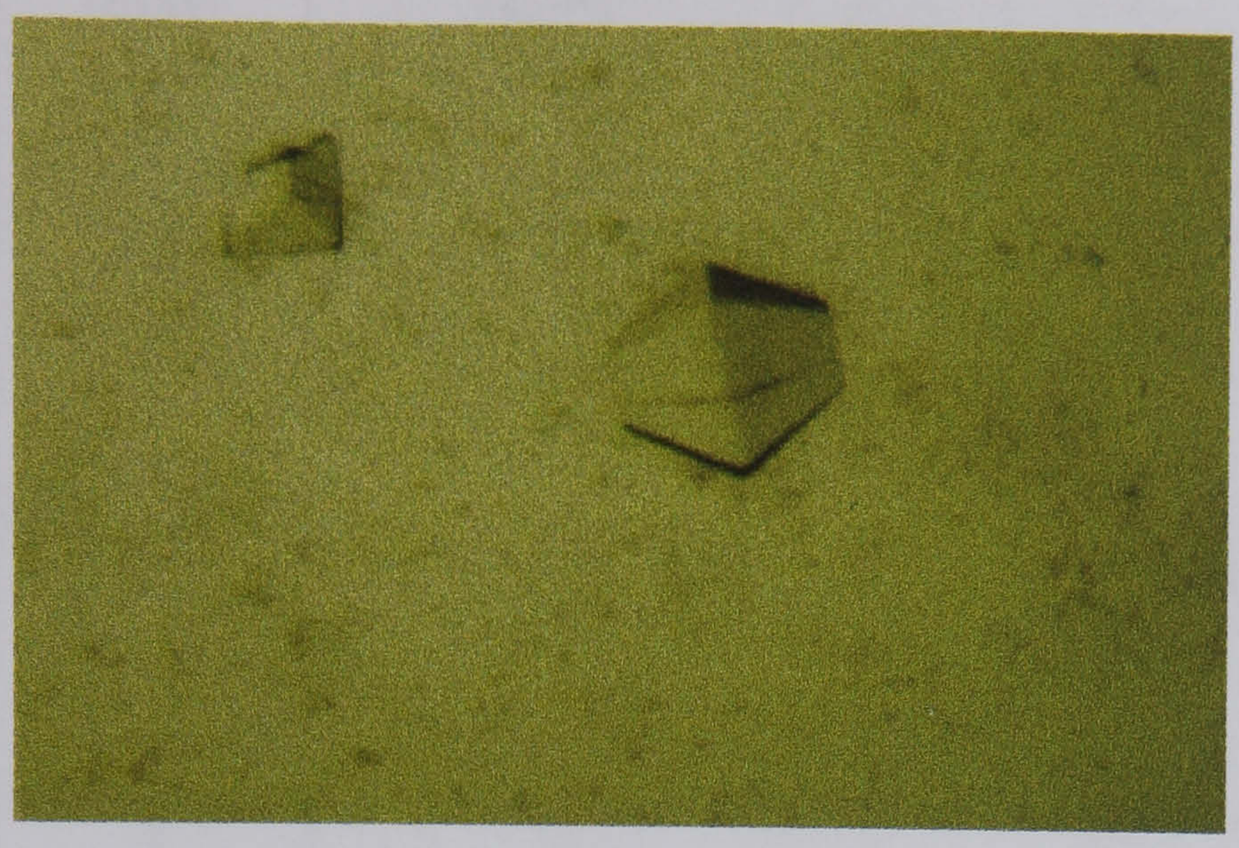


Figure 4.14 Protein crystals grown using ternary complex crystallisation condition (1).

Ternary complex Crystal form (1)

These crystals diffracted poorly (~5.5Å), therefore data were not collected.

One further condition for crystallising the ternary complex was found using Hampton Screen (I).

Ternary complex crystallisation condition (2): 28mg/ml protein, 4mM tNADH, 25mM α-ketoisocaproate in the 1.5µl + 1.5µl drop solution; 8% PEG 4000, 0.1M Na acetate pH 4.6 in the well solution (Table 4.13).

[Protein] (mg/ml)	[tNADH] (mM)	[α-ketoisocaproate] (mM)	[PEG 4000] Range (%)	pH Range	Results (after ~2 weeks)
28 (1.5μl+1.5μl drop)	4	25	7 - 12	4.4-5.0	Small cubic shaped crystals found in drops initially containing 9-10% PEG.*
28	4	25	7 - 12	4.4-5.0	Cubic crystals found in several drops (~400μm^3). Other drops contain precipitate.*
10	4	25	7 - 12	4.4-5.0	Drops contain light precipitate.
10	4	5	4 - 16	4.0-5.0	Drops contain light precipitate.
15	1	5	4 - 16	4.0-5.0	Most drops contain some light precipitate.
10	1	5	18 - 24	5.5-7.0	All drops contain heavy precipitate.
15	1	5	18 - 24	5.5-7.0	All drops contain heavy precipitate.
15	2	20	3 - 7	4.0-5.0	Drops remain clear.

Table 4.13 A selection of trials used to optimise ternary complex crystallisation condition (2).

The crystals found in the first two trays in table 4.8 diffracted well; very many attempts were made to repeat these crystallisation conditions: Unfortunately, heavy precipitate developed in most of the drops. It is still unclear as to why these crystallisation conditions (in which by far the best H205Q D-HicDH crystals grew) could not be reproduced.

Ternary complex Crystal form (2)

Data collected at the Daresbury SRS laboratory (PX7.2)

Cryoprotectant solution	15% PEG 4000, 0.1M Na acetate pH 4.5, 4mM tNADH, 25mM α-ketoisocaproate, 30% glycerol
Unit cell	a=80.11Å, b=80.12Å, c=277.24Å, α=ß=γ=90°
Space group	$P4_1$, $P4_3$, $P4_122$ or $P4_322$
Total number of unique reflections	13,336
Mean Redundancy	2.9
Resolution	2.6Å (81.1% complete) 2.9Å (96.6% complete)
Mosaicity	0.57°
Completeness - Overall (19.96-2.9Å) Outer shell (3.05-2.9Å)	96.6% 91.9%
I / σ (I) - Overall Outer shell	20.3 4.4
R_{merge} - Overall Outer shell	8.7% 15.2%

Table 4.114 Diffraction data collection statistics of a ternary complex crystallisation condition (2) crystal.

The crystal diffracted beyond 2.4Å, but there were many reflections overlaps due to the very long unit cell c axis. Therefore the detector had had to be pushed back to a distance where the resolution at the edge of the image plate was 2.6Å, and the oscillation range reduced from 1.5° to 0.8°. However, still there were too

many reflection overlaps and so diffraction data was collected to a resolution of 2.9Å using a 1° oscillation range.

The data were processed with primitive tetragonal symmetry in *Denzo*. The data were scaled in *Scalepack* using the space groups $P4_1$ and $P4_3$, this showed that there were systematic absences at the refection conditions l = 4n. However, there were no systematic absences at l = 2n; therefore the space groups could not $P4_2$, $P4_222$ or P4 , assuming that the crystal did not exhibit pseudosymmetry in space group P4, as these space group show reflection absences at l = 2n, 6n, 10n, etc. (Hahn, 1995). Also, there were no absences at k = 2n, so the space groups could not be $P4_32_12$ and $P4_12_12$, as these space groups with a two fold axes have reflections absent at k = 2n. Thus the space groups could be $P4_1$ or $P4_3$, or their higher space groups $P4_322$ or $P4_122$ which do not have two fold screw axes. The two sets of space group had approximately the same R_{merge} values, so this did not help in determining the space group of the crystal.

Chapter 5

Attempts at Phasing the D-2-Hydroxyisocaproate Dehydrogenase Diffraction Data Sets

5.1 Molecular Replacement

5.1.1 Choosing the Search Models

The Protein Data Bank (PDB) (Berman *et al.*, 2000) is the single worldwide archive structural database of biological macromolecules and is thus the most suitable place to search for proteins with known structure which have sequence (and therefore presumably structural) similarity with the target protein. For the molecular replacement phasing method, search models usually have to show at least 20% sequence similarity (Tickle, 1998).

The European Bioinformatics Institute's Macromolecular Structure Database (*EBI-MSD*) program (Keller *et al.*, 1998) was used to search the PDB with the protein sequence of *lactobacillus delbrueckii* subsp. *bulgaricus* D-hydroxyisocaproate dehydrogenase (accession I.D.: Q48534) (SWISS-PROT protein sequence database; Bairoch A. and Apweiler R. 2000) (Table 5.1).

PDB I.D.	Source and Name of Protein	Sequence identity (%)[a]
-	*Lactobacillus bulgaricus* D-2-Hydroxyisocaproate Dehydrogenase	100
2DLD	*Lactobacillus helveticus* D-Lactate Dehydrogenase	38
1DLD	*Lactobacillus bulgaricus* D-Lactate Dehydrogenase	36
1DXY	*Lactobacillus casei* D-2-Hydroxyisocaproate Dehydrogenase	34
1PSD	*Escherichia coli* D-3-Phosphoglycerate Dehydrogenase	30
1GDH	*Hyphomicrobium methylovorum* D-Glycerate Dehydrogenase	28
2NAD	*Pseudomonas* Sp.101 Formate Dehydrogenase	28
1QP8	*Pyrobaculum Aerophilum* Formate Dehydrogenase	25
1PGJ	*Trypanosoma brucei* 6-Phosphogluconate Dehydrogenase	21

[a] Sequence identity over the entire 333 amino acid residues of *Lactobacillus bulgaricus* D-2-Hydroxyisocaproate Dehydrogenase.

Table 5.1 The protein structures found in the PDB which can be used as search models in the molecular replacement phasing of D-HicDH.

5.1.2 Estimating the Number of Molecules in the Asymmetric Unit

In order to determine the number of rotation and translation solutions expected during the molecular replacement procedure, the number of molecules per asymmetric unit must be known : This was estimated using the method proposed by Matthews (1968).

$$V_M = \frac{cellvolume}{Mn_{asymu}nmols_{asu}} = \frac{V}{MZ}$$

where:-
M is the molecular weight of the protein in Daltons (computed from the primary sequence using the *ExPASy-tools* programs, 2001)
V_M is the Matthews coefficient
V is the cell volume ($Å^3$)
n_{asymu} is the number of asymmetric units, a.s.u., (determined by the space group of the protein crystal)
$nmols_{asu}$ is the number of molecules in the asymmetric unit
Z is the number of molecules in the unit cell.

For many protein crystals, V_M is usually between 1.66 and 4.0 $Å^3$/Da, corresponding to protein contents of 30 - 75%. Most values of V_M are around 2.5 $Å^3$/Da (~50% solvent in the crystal). This helps to narrow down the possibilities for $nmols_{asu}$.

V_M values were determined using the *Matthews_coef* CCP4 program.

5.1.3 Attempts at Solving the D-HicDH Structure using *AMoRe*

5.1.3.1 The *AMoRe* Program

The *AMoRe* (Automated Molecular Replacement) program package (Navaza, 1994) (CCP4) was initially used in attempting to solve the structure of D-HicDH.

There are several operations required for structure solution using the *AMoRe* program.

5.1.3.2 Sorting and Tabling

The sorting operation converted the truncated data reflection file into a format used by the *AMoRe* program.

After removing the PDB header information, solvent molecules and any ligands from the search model's structural file, the tabling operation was applied. This rotated the model so that its centre of mass was parallel to the coordinate axes, and the centroid of the coordinates was shifted to the origin, this helps to minimise the cell size required in the rotation operation (Wolf *et al.*, 1991). Tabling also computed a molecular Fourier transform for use in structure factor calculations later.

The search model was then viewed in the *Quanta* graphics program, and an integration radius of 75% of the minimum molecular radius was chosen to ensure that no cross molecular vectors were included in the Patterson search, as recommended by Blow (1995).

Sorting and Tabling provided a maximal distance from the centre of mass which was needed to calculate the size of the P1 unit cell used in the rotation function. The cell had to be at least the box bounding the model plus the chosen radius in each direction, this ensured that all the calculated inter-molecular Patterson vectors were excluded. But, the cell could not be too large otherwise too many structure factors would have been calculated (Navaza, 1994). Dodson (1995) has suggested that the cell dimensions should be twice the maximal distance from the centre of mass plus plus the chosen radius plus a small safety term (5Å was used). The unit cell had angles all at 90°.

5.1.3.3 The Rotation Function

During the rotation function, structure factor amplitudes were calculated for the search model in the P1 cell. Spherical harmonic coefficients were then calculated for the target and search model Patterson maps. A cross-rotation function was then calculated. This finds possible solutions by Patterson overlap. There should be as many peaks standing out above the noise as there are molecules in the asymmetric unit.

The radius of integration, unit cell dimensions and resolution cutoffs were then altered to obtain the best solution.

5.1.3.4 The Translation Function and Fitting

The best 10-30 cross-rotation solutions were then put into the translation function in order to compute the translation solutions. In order to improve the subsequent solution(s) the resolution cutoffs were changed.

If there is more than one molecule in the asymmetric unit, then the first solution was fixed and a translation solution computed for the second molecule, this procedure was then repeated for the other remaining molecules.

Rigid body refinement (i.e. the fitting function) was then carried out on the best translation solution(s).

5.1.3.5 Attempts at Phasing the D-HicDH Datasets

(i) Using the H205Q D-HicDH ternary complex diffraction data obtained from crystals grown using condition 2 (in Chapter 4)

The data was 96.6% complete at 2.9Å and 81.1% complete at 2.6Å. When processed in $P4_1$ or $P4_3$: $V_M = 12.4$ Å^3/Da $\Rightarrow$ 4 molecules/a.s.u.

When processed in $P4_122$ or $P4_322$: $V_M = 6.2$ Å^3/Da $\Rightarrow$ 2 molecules/a.s.u.

Many attempts were made at finding rotation solutions using the data processed in the various space groups and using all the different search models, with numerous different *AMoRe* parameters. This included employing 'modified' search models such as:-

(1) Cα traces of the search models.

(2) Monomers of the models with flexible loops and termini removed.

(3) Just the structurally conserved NAD^+-binding regions.

(4) Poly-alanine, poly-glycine and poly-serine traces, produced using *Moleman2* (part of the Uppsala suite of programs; Kleywegt, 2000).

(5) Superimposition of some of the search models, using the CCP4 program *LSQKAB* (which minimises the root mean square deviations between specified regions of the models). This was done to enhance regions of structural similarity (Tickle, 1998).

(6) Combinations of the above models.

In nearly all cases there were no clear rotation solutions. However, small rotation peaks above the noise were observed when using the data processed in $P4_122$ and $P4_322$ and a monomer of *L.casei* D-HicDH (Dengler *et al.*, 1997) as the search model. The space groups $P4_122$ and $P4_322$ should, and indeed did, possess the same rotation functions but different translation functions. The best of these rotation solutions are shown in Table 5.1. Only one very small rotation peak above the noise was observed in this (and all subsequent) cases.

Search model: monomer of L.casei D-HicDH at 1.9 Å.

Radius of integration: 25Å. Cell model: 105Å × 105Å × 105Å. Resolution cutoff: 10.0-3.5Å.

Solution Number	**Rotation of Eulerian angles (°)**			**Correlation coefficient (%)**
	α	β	γ	
Solution 1	74.44	60.47	84.87	10.8
Solution 2	81.24	61.56	44.45	8.4
Solution 3	63.41	56.73	207.34	8.0
Solution 4	39.81	86.94	277.43	7.9
Solution 5	23.70	41.32	224.54	7.8
Solution 6	87.68	83.09	172.55	7.4
Solution 7	51.92	65.24	176.05	7.4
Solution 8	84.55	81.80	56.97	7.4
Solution 9	11.18	55.89	204.78	7.2
Solution 10	30.95	61.41	72.12	7.2

Table 5.2 The rotation solutions from *AMoRe* for the ternary H205Q D-HicDH complex processed in the $P4_122$ and $P4_322$ space groups using the monomer of *L.casei* D-HicDH as the search model.

The top 30 rotation solutions were then put into the translation function. However, even after altering the various *AMoRe* parameters no peak above the noise in the translation function was observed in either the $P4_122$ (Table 5.3) or $P4_322$ (Table 5.4) forms. Furthermore, the rotation solutions appeared to be consistent with a random array of atoms. Therefore attempts to improve the R-factor by searching for the remaining molecules in the asymmetric unit failed.

Solution Number	**Rotation of Eulerian angles (°)**			**Translation (fractions of the unit cell)**			**Correlation coefficient**	**R-factor**
	α	β	γ	**x**	**y**	**z**	(%)	(%)
Solution 1	74.44	60.47	84.87	0.9289	0.1938	0.1060	8.3	57.3
Solution 2	81.24	61.56	44.45	0.1465	0.1632	0.3854	9.1	57.3
Solution 3	63.41	56.73	207.34	0.0994	0.2556	0.3937	9.8	56.9
Solution 4	39.81	86.94	277.43	0.7224	0.5916	0.3535	8.8	57.6
Solution 5	23.70	41.32	224.54	0.3158	0.4544	0.4846	8.9	57.3
Solution 6	87.68	83.09	172.55	0.9453	0.6131	0.0592	9.9	57.6
Solution 7	51.92	65.24	176.05	0.3539	0.4831	0.0937	9.3	57.4
Solution 8	84.55	81.80	56.97	0.9452	0.6222	0.2587	9.8	57.3
Solution 9	11.18	55.89	204.78	0.9419	0.8875	0.3705	9.4	57.5
Solution 10	30.95	61.41	72.12	0.0787	0.0747	0.3885	8.3	57.6

Table 5.3 The translation solutions from *AMoRe* for the ternary H205Q D-HicDH complex processed in the $P4_122$ space group using the monomer of *L.casei* D-HicDH as the search model.

Solution Number	**Rotation of Eulerian angles (°)**			**Translation (fractions of the unit cell)**			**Correlation coefficient**	**R-factor**
	α	β	γ	**x**	**y**	**z**	(%)	(%)
Solution 1	74.44	60.47	84.87	0.1367	0.5498	0.4567	9.3	57.6
Solution 2	81.24	61.56	44.45	0.7856	0.0987	0.6767	9.6	57.7
Solution 3	63.41	56.73	207.34	0.6745	0.3351	0.2156	9.7	57.7
Solution 4	39.81	86.94	277.43	0.2365	0.8734	0.1568	9.4	58.2
Solution 5	23.70	41.32	224.54	0.5765	0.4509	0.2678	9.6	57.9
Solution 6	87.68	83.09	172.55	0.0957	0.7567	0.4537	9.3	58.3
Solution 7	51.92	65.24	176.05	0.591	0.6433	0.6876	8.5	58.3
Solution 8	84.55	81.80	56.97	0.0837	0.2547	0.6232	9.1	58.4
Solution 9	11.18	55.89	204.78	0.9945	0.4556	0.7845	8.7	58.5
Solution 10	30.95	61.41	72.12	0.0653	0.3587	0.1432	8.9	58.5

Table 5.4 The translation solutions from *AMoRe* for the ternary H205Q D-HicDH complex processed in the $P4_322$ space group using the monomer of *L.casei* D-HicDH as the search model.

(ii) Using the wild type D-HicDH binary complex diffraction data obtained from crystals grown using condition 1

The data was 84.1% complete at 3.1Å. V_M = 8.9 Å^3/Da $\Rightarrow$ 4 molecules/a.s.u.

Numerous attempts were made at finding a rotation solution; the best rotation functions were found when using the NAD^+-binding domain monomer of *L.casei* D-HicDH (the ~150 amino acid residue model prepared after an examination of the structure in *Quanta*) (Table 5.5).

Search model: NAD^+-binding domain monomer of L.casei D-HicDH at 1.9Å.

Radius of integration: 19Å. Cell model: 92Å × 92Å × 92Å. Resolution cutoff: 9.0-4.1Å.

Solution Number	**Rotation of Eulerian angles (°)**			**Correlation coefficient**
	α	β	γ	(%)
Solution 1	227.94	91.48	90.15	11.3
Solution 2	18.91	113.74	358.11	10.6
Solution 3	70.18	65.23	37.95	10.5
Solution 4	198.70	88.13	12.00	10.4
Solution 5	91.00	24.67	44.00	10.4
Solution 6	56.61	76.01	29.93	10.0
Solution 7	181.28	110.26	223.66	9.9
Solution 8	161.40	68.16	170.09	9.8
Solution 9	336.50	23.12	206.00	9.8
Solution 10	280.56	100.93	342.83	9.7

Table 5.5 The rotation solutions from *AMoRe* for the binary wild type D-HicDH complex using the NAD^+-binding domain monomer of *L.casei* D-HicDH as the search model.

The top 20 rotation solutions were then put into the translation function. Again, the translation function results were consistent with a random array of atoms (Table 5.6).

Solution Number	Rotation of Eulerian angles (°)			Translation (fractions of the unit cell)			Correlation coefficient	R-factor
	α	β	γ	x	y	z	(%)	(%)
Solution 1	227.94	91.48	90.15	0.1711	0.9923	0.3765	41.3	57.6
Solution 2	18.91	113.74	358.11	0.1970	0.0000	0.4164	35.8	56.8
Solution 3	70.18	65.23	37.95	0.0917	0.0705	0.3645	40.6	57.5
Solution 4	198.70	88.13	12.00	0.1970	0.0000	0.4164	35.8	56.8
Solution 5	91.00	24.67	44.00	0.1010	0.9141	0.4574	40.6	57.6
Solution 6	56.61	76.01	29.93	0.1970	0.0000	0.4164	35.8	56.8
Solution 7	181.28	110.26	223.66	0.7587	0.5083	0.6657	40.0	57.7
Solution 8	161.40	68.16	170.09	0.1970	0.0000	0.4164	35.8	56.8
Solution 9	336.50	23.12	206.00	0.8295	0.4926	0.6238	41.1	57.6
Solution 10	280.56	100.93	342.83	0.1970	0.0000	0.4164	35.8	56.8

Table 5.6 The translation solutions from *AMoRe* for the binary wild type D-HicDH complex using the NAD^+-binding domain monomer of *L.casei* D-HicDH as the search model.

The inability to solve the wild type binary complex using MR may have been due to the incompleteness of the observed data (84.1% complete). There is considerable evidence to suggest that data incompleteness can have a negative effect on the solution of a structure by molecular replacement (Dodson *et al.*, 1992; Dodson, 1995).

(iii) Using the H205Q D-HicDH binary complex diffraction data obtained from crystals grown using condition 3a ($P2_1$)

The data was 91.6% complete at 3.0Å. V_M = 9.4 $Å^3$/Da $\Rightarrow$ 4 molecules/a.s.u.

Small rotation peaks above the noise were found using *e.g.* a ~240 amino acid residue poly-alanine trace of the NAD^+-binding domain monomer of *L.bulgaricus* D-LDH (Vinals *et al.*, 1995) (a knowledge-based modelled structure).

Search model: Poly-alanine trace of the NAD^+-binding domain monomer of L.Bulgaricus D-LDH.

Radius of integration: 25Å. Cell model: 100Å × 100Å × 100Å. Resolution cutoff: 8.0-4.0Å. (Few reflections in this resolution range)

Solution Number	**Rotation of Eulerian angles (°)**			**Correlation coefficient**
	α	β	γ	(%)
Solution 1	258.55	79.22	209.77	11.9
Solution 2	353.50	52.23	93.03	10.7
Solution 3	335.73	69.70	212.38	10.6
Solution 4	121.50	35.79	91.53	10.4
Solution 5	242.38	90.00	258.22	10.3
Solution 6	297.77	90.00	77.56	10.3
Solution 7	328.40	76.28	211.45	10.3
Solution 8	126.25	77.47	209.20	10.2
Solution 9	209.50	90.00	111.09	10.0
Solution 10	329.26	90.00	291.40	10.0

Table 5.7 The rotation solutions from *AMoRe* for the binary H205Q D-HicDH complex using the poly-alanine trace of the NAD^+-binding domain monomer of *L.bulgaricus* D-LDH as the search model.

Solution Number	**Rotation of Eulerian angles (°)**			**Translation (fractions of the unit cell)**			**Correlation coefficient**	**R-factor**
	α	β	γ	**x**	**y**	**z**	(%)	(%)
Solution 1	258.55	79.22	209.77	0.1000	0.0000	0.4762	10.4	57.6
Solution 2	353.50	52.23	93.03	0.2250	0.0000	0.2619	10.2	58.2
Solution 3	335.73	69.70	212.38	0.2000	0.0000	0.0595	8.9	58.4
Solution 4	121.50	35.79	91.53	0.2250	0.0000	0.1310	12.8	57.2
Solution 5	242.38	90.00	258.22	0.1125	0.0000	0.4167	11.5	57.3
Solution 6	297.77	90.00	77.56	0.0250	0.0000	0.0238	11.5	57.4
Solution 7	328.40	76.28	211.45	0.2000	0.0000	0.0595	8.8	58.1
Solution 8	126.25	77.47	209.20	0.0125	0.0000	0.1310	9.4	57.9
Solution 9	209.50	90.00	111.09	0.2625	0.0000	0.0595	11.1	57.2
Solution 10	329.26	90.00	291.40	0.2375	0.0000	0.4405	11.2	57.2

Table 5.8 The translation solutions from *AMoRe* for the binary H205Q D-HicDH complex using the poly-alanine trace of the NAD^+-binding domain monomer of *L.bulgaricus* D-LDH as the search model.

(iv) Using the H205Q D-HicDH binary complex diffraction data obtained from crystals grown using condition 3b (C2)

The data was 95.2% complete at 2.8Å. V_M = 9.6 Å^3/Da $\Rightarrow$ 4 molecules/a.s.u.

Rotation peaks were observed, usually with search model dimers (Table 5.7), but again there were no convincing translation solutions (Table5.8).

Search model: Dimer of L.Helveticus D-LDH (Dunn and Holbrook, 1996) at 2.7 Å.

Radius of integration: 34Å. Cell model: 125Å × 125Å × 125Å. Resolution cutoff: 15.0-3.5Å.

Solution Number	**Rotation of Eulerian angles (°)**			**Correlation coefficient (%)**
	α	β	γ	
Solution 1	0.37	78.01	122.46	11.3
Solution 2	241.00	44.83	178.82	8.3
Solution 3	86.44	45.87	160.47	8.3
Solution 4	267.87	13.57	38.43	8.0
Solution 5	356.77	85.01	340.07	7.8
Solution 6	335.02	49.21	27.66	7.5
Solution 7	349.78	41.95	258.47	7.5
Solution 8	41.00	12.42	84.00	7.3
Solution 9	121.18	72.80	278.16	7.3
Solution 10	330.31	54.79	133.26	7.3

Table 5.9 The rotation solutions from *AMoRe* for the binary H205Q D-HicDH complex using the dimer of *L.helveticus* D-LDH as the search model.

Solution Number	Rotation of Eulerian angles (°)			Translation (fractions of the unit cell)			Correlation coefficient	R-factor
	α	β	γ	x	y	z	(%)	(%)
Solution 1	0.37	78.01	122.46	0.2708	0.0000	0.2115	39.2	55.9
Solution 2	241.00	44.83	178.82	0.3542	0.0000	0.3462	36.4	57.1
Solution 3	86.44	45.87	160.47	0.1042	0.0000	0.2500	33.2	56.8
Solution 4	267.87	13.57	38.43	0.2083	0.0000	0.4423	31.4	57.2
Solution 5	356.77	85.01	340.07	0.3958	0.0000	0.4808	42.4	56.5
Solution 6	335.02	49.21	27.66	0.3333	0.0000	0.0000	42.8	55.3
Solution 7	349.78	41.95	258.47	0.1667	0.0000	0.0000	39.0	56.3
Solution 8	41.00	12.42	84.00	0.2708	0.0000	0.3077	38.4	55.7
Solution 9	121.18	72.80	278.16	0.2917	0.0000	0.0385	40.1	57.1
Solution 10	330.31	54.79	133.26	0.3750	0.0000	0.0192	40.4	55.9

Table 5.10 The translation solutions from *AMoRe* for the binary H205Q D-HicDH complex using the dimer of *L.helveticus* D-LDH as the search model.

The inability to find a solution may have again been to due to the incompleteness of the observed data. Similar problems were encountered when trying to solve the structure from the remaining D-HicDH incomplete low resolution data sets.

5.1.4 Attempts at Solving the D-HicDH Structure using *AMoRe* with Normalised Structure Factor Amplitudes

5.1.4.1 Normalised Structure Factor Amplitudes

Normalised structure factor amplitudes (*E*'s) are structure factor amplitudes calculated from diffraction data that has been divided up into resolution shells of constant volume and then scaled so that the mean intensity does not decrease with resolution. The scaling usually involves applying an exponential 'temperature' factor to the structure factors (Dodson *et al.*, 1992). The contrast of the Patterson map and the peaks should be better resolved if the data has been normalised (Dodson *et al.*, 1992; Tickle, 1998).

The calculation of rotation and translation functions in *AMoRe* using *E*'s required the use of several CCP4 programs.

5.1.4.2 Calculating *E* values, Sorting and Tabling, and the Rotation Function

After removing the header information from the search model's PDB file, structure factor amplitudes for the model in a P1 cell were calculated using *SFALL*. The observed and calculated (search model) structure factor amplitudes were then normalised using the *ECALC* program.

The sorting, tabling and cross-rotation functions were then carried out in *AMoRe* as usual.

5.1.4.3 The Translation Function using *E* Values

PDBSET was used to apply the rotation matrix to the search molecule (a rotation matrix for each molecule in the asymmetric unit should have been applied. However, only one rotation peak above the noise was ever found). Structure factor amplitudes and phases were then calculated using *SFALL*.

CAD was used then used to combine the observed and calculated data into a single file. Fourier coefficients were calculated for the translation function using *TFFC* and then *FFT* (Fast Fourier Transform) used to obtain an intra-molecular vector translation function map. The signal to noise ratios (S / σ) of the subsequent peaks were then studied.

5.1.4.4 Attempts at Phasing the D-HicDH Datasets using *E*'s

Unfortunately no improvement in the solutions was obtained with *AMoRe* when using normalised structure factor amplitudes. For example:

Using the H205Q D-HicDH binary complex diffraction data obtained from crystals grown using condition 3a ($P2_1$)

The data was 94.6% complete at 3.0Å. V_M = 9.4 Å^3/Da $\Rightarrow$ 4 molecules/a.s.u.

Search model: Poly-alanine dimer of H.Methylovorum D-GDH at 2.4 Å.

Radius of integration: 36Å. Resolution cutoff: 10.0-4.5Å.

Solution Number	**Rotation of Eulerian angles (°)**			**Correlation coefficient (%)**
	α	β	γ	
Solution 1	298.52	65.94	83.57	8.8
Solution 2	129.00	56.38	101.00	8.2
Solution 3	24.17	69.16	39.10	8.1
Solution 4	28.50	20.28	139.00	7.9
Solution 5	291.00	46.19	106.50	7.2
Solution 6	135.32	65.51	71.99	6.8
Solution 7	271.09	77.33	141.71	6.6
Solution 8	40.66	71.15	261.31	6.6
Solution 9	327.14	72.55	225.92	6.3
Solution 10	223.91	47.06	176.81	6.1

Table 5.11 The rotation solutions from *AMoRe* using normalised structure factor amplitudes for the binary H205Q D-HicDH complex, with the poly-alanine dimer of *H.methylovorum* D-GDH as the search model.

Solution Number	**Translation (fractions of the unit cell)**			S/σ	$\Delta S/\sigma$
	t_x	t_y	t_z		
Solution 1	0.277	0.002	0.271	6.02	0.61
Solution 2	0.316	0.000	0.319	5.41	-

Table 5.12 The translation function map peaks using normalised structure factor amplitudes from *TFFC* for the binary H205Q D-HicDH complex, with the poly-alanine dimer of *H.methylovorum* D-GDH as the search model.

A ΔS / σ ratio greater than 1 usually indicates a correct solution (Tickle, 1998). Therefore the solution above is unlikely to be correct.

5.1.5 Attempts at Solving the D-HicDH Structure using *EPMR*

EPMR is a molecular replacement program that utilises an evolutionary search algorithm (Kissinger *et al.*, 1999). It is a highly automated program that requires three input files, the search model PDB file; a file containing the crystal's space group and unit cell dimensions; and finally a file containing just the *hkl* indices and the corresponding observed structure factor amplitudes.

The six-dimensional search carried out by the program first involves generating an initial set of random orientations and positions for the search model. A Patterson correlation coefficient is then calculated for each orientation, and the highest scoring orientations are retained and used to generate a new set of trial orientations. This is achieved by applying random alterations to the orientation angles and translations for each 'parent' orientation. This procedure is then repeated for a specified number of cycles.

At the end of the evolutionary optimisation rigid-body refinement is then carried out on the search model.

As well as allowing a specified number of runs of the evolutionary optimisation process, *EPMR* also allowed the search for multiple molecules in the asymmetric unit. The first solution was kept as a partial structure and the PDB file written out, whilst the program kept looking for the remaining solutions.

Before starting the program, variables such as the resolution cutoffs, number of molecules to be located, number of optimisation cycles, number of runs, *etc.* were entered. Unfortunately, even when executing '50 runs' (about 4 days of computing time), *EPMR* still failed to find any sort of

molecular replacement solution. Moreover, the subsequent correlation coefficients and R-factors were consistent with a random array of atoms.

Three examples Using the H205Q D-HicDH binary complex diffraction data obtained from crystals grown using condition 3a ($P2_1$)are shown below:-

(The data was 94.6% complete at 3.0Å with an estimated 4 molecules in the a.s.u).

(a) *Search model: Monomer of L.Bulgaricus D-LDH.*

Number of Solutions: 4. Number of cycles: 50. Number of Runs: 50.

Resolution cutoffs: 15.0 - 4.0Å.

	Rotation of Eulerian angles (°)			**Translation (fractions of the unit cell)**			**Correlation coefficient**	**R-factor**
	α	β	γ	**x**	**y**	**z**	(%)	(%)
Final Solution	226.39	75.60	3.36	0.7110	0.9828	0.7389	42.5	58.0

(b) *Search model: NAD^+-binding domain monomer of L.casei D-HicDH.*

Number of Solutions: 1. Number of cycles: 50. Number of Runs: 20.

Resolution cutoffs: 15.0 - 4.0Å.

	Rotation of Eulerian angles (°)			**Translation (fractions of the unit cell)**			**Correlation coefficient**	**R-factor**
	α	β	γ	**x**	**y**	**z**	(%)	(%)
Final Solution	58.39	164.02	2.15	0.6120	0.4449	0.1100	25.3	61.0

(c) *Search model: Poly-alanine dimer of H.Methylovorum D-GDH.*

Number of Solutions: 4. Number of cycles: 50. Number of Runs: 20.

Resolution cutoffs: 15.0 - 4.0Å.

	Rotation of Eulerian angles (°)			**Translation (fractions of the unit cell)**			**Correlation coefficient**	**R-factor**
	α	β	γ	**x**	**y**	**z**	(%)	(%)
Final Solution	312.07	75.01	40.35	0.3822	0.5277	0.3149	42.2	57.6

Table 5.13 The rotation and translation solutions from *EPMR* for the binary H205Q D-HicDH complex using a variety of search models.

5.1.6 Conclusions

The *XPLOR* program (Brünger, 1992a) was also used to try and solve the best D-HicDH data sets. However, even though this program applied Patterson correlation refinement (*i.e.* rigid-body refinement in Patterson space) to the computed cross-rotation functions, no rotation peaks above the noise were ever observed.

The inability to solve the D-HicDH structure with some of the data sets using molecular replacement may have been down to the general poor quailty and incompleteness of the data.

Curiously, attempts to solve the better diffraction data sets with the structurally conserved NAD^+-binding domain as the search model also failed. One possible explanation for this is that there may be substantial structural differences in the cofactor-binding domains of the target and search molecules: Chothia and Lesk, (1986) showed that the same general fold differs in structure by amounts that increase as the amino acid sequence diverge. After comparing the structures of a number of proteins Chothia and Lesk concluded that the divergence in structure of two homologous cores could be measured as a simple function of the fractional identity of the cores:

$$\Delta = 0.40e^{1.87H}$$

where
Δ = the root mean square deviation (r.m.s.d.) in the main chain atoms.
H = the fraction of mutated residues in the homologous region.

Therefore if the sequence identity is > 50% the protein structure should provide a close general model for the unknown protein structure. However, if the amino acid sequence identity between the two proteins is about 20% there will probably be large structural differences.

It was shown that differences in the common cores consist mainly of changes in the relative position and orientation of packed secondary structures. Moreover, in the case of β-sheets (found in *e.g.* the cofactor-binding domains) local changes in structure were also observed.

5.2 Multiple Isomorphous Replacement

As attempts at solving the D-HicDH structures using molecular replacement were not successful, the multiple isomorphous replacement method was attempted.

5.2.1 Heavy Atom Soaking Procedures

The methods employed in the soaking of crystals prior to cryogenic data collection involved gradually soaking the protein crystals in mother liquor containing increasing concentrations of the heavy atom salt and glycerol (often with a slightly increased precipitant concentration).

Preliminary investigations were carried out to determine the maximum duration for a particular heavy atom soak and the optimum concentrations of the salt needed. This was achieved by observing soaked crystals at regular time intervals and noting when cracks started to appear, morphology started to change, and / or the protein crystals started to lose their birefrigence.

5.2.2 The Heavy Atom Salts Utilised

Many different heavy atoms salts were employed in the search for useful derivative crystals for MIR phasing. However, the majority of the salts utilised caused a severe deterioration in the quailty of the crystals resulting in very poor diffraction and / or the inability to collect complete derivative data sets. This problem was compounded by the fact that the corresponding native crystals were themselves weakly diffracting.

It was found that the H205Q D-HicDH binary complex crystals grown using condition (3a) in Chapter 4 (*i.e.* 35-40mg/ml protein, 1mM NADH; 20-25% PEG MME 5000, 0.2M ammonium sulphate, 0.1M MES pH 6.5), were the most suitable for heavy atom soaking studies.

Derivative number	Crystal soaking conditions	Observed maximum resolution (Å)
1	24 hours in 1mM uranyl acetate	~5.5
2	6 days in 1mM uranyl acetate (very large crystal)	3.5
3	19 hours in 1mM platinum potassium nitrate	3.2
4	5 hours in 1mM p-chloromercuribenzene sulphonic acid	5.8
5	5 hours in 0.5mM p-chloromercuribenzene sulphonic acid	5.6
6	12 hours in 2mM mercuric acetate	5.2
7	48 hours in 2mM mercuric cyanide	4.9
8	48 hours in 2mM platinum potassium nitrate	2.9
9	24 hours in 2mM cadmium lactate	5.8
10	15 hours in 2mM thallous acetate	4.0

Derivative number	Crystal soaking conditions	Observed maximum resolution (Å)
11	24 hours in 8mM uranyl nitrate	3.3
12	48 hours in 4mM uranyl nitrate	~5.5
13	20 hours in 2mM platinum potassium chloride	~6.0
14	15 hours in 0.5mM platinum potassium chloride	4.9
15	1 hour in 0.5mM potassium uranyl iodate	~5.5
16	3 hours in 4mM samarium chloride	3.2
17	24 hours in 1mM samarium chloride	5.2
18	24 hours in 3mM iridium chloride	5.1
19	12 hours in 4mM cadmium acetate	~5.5
20	16 hours in 3mM erbium chloride	~5.0

Table 5.14 A selection of X-ray diffraction tests on various heavy atom soaked crystals.

The X-ray diffraction was found to be even worse when using the other crystals forms of H205Q D-HicDH in heavy atom soaking experiments.

5.2.3 Diffraction Data Collection of Potential Derivatives

Only a handful of the numerous soaked crystals diffracted to a sufficiently high resolution for attempts to be made at collecting a complete data set. However, most of those chosen had a high mosaicity. The *Strategy* computer program was always utilised in order to calculate an optimised data collection strategy.

Derivative ⇒ number	2 (uranyl acetate)	3 (platinum potassium nitrate)	8 (platinum potassium nitrate)	10 (thallous acetate)	11 (uranyl nitrate)	17 (samarium chloride)
Number of crystals	1	2	1	3	2	1
Total number of observations	36,283	17,986	25,616	12,085	10,499	42,577
Redundancy	2.9	1.6	2.5	2.6	2.2	2.8
Resolution in the outer shell (Å)	4.0-3.6	4.3-4.0	4.1-4.0	5.0-4.8	4.7-4.4	3.8-3.2
Mosaicity	1.17°	0.86°	1.29°	0.86°	1.68°	1.60°
Mean I / σ (I) in the outer resolution shell	2.96	4.14	3.66	2.56	3.14	3.93
Completeness in the outer resolution shell (%)	88.7	89.4	82.1	69.4	51.8	93.4
Overall R_{merge} (%)	22.6	23.0	38.1	32.2	21.4	22.0

Table 5.15 Diffraction data collection statistics of heavy atom soaked crystals.

5.2.4 Location and Refinement of possible Heavy Atom Sites

Once a putative derivative data set had been collected and reduced, a series of CCP4 programs were used in order to locate and refine the position(s) of any possible heavy atom binding sites.

5.2.4.1 Scaling the Derivative and Native Datasets

Initially, the computer program *CAD* was used to combine and sort the truncated native and putative derivative reflection data into one file.

The X-ray diffraction experiment does not take into account variations in the incident beam intensity, wavelength, crystal size and all other factors. Therefore, an absolute scale-factor has to be calculated to put the and native and derivative structure factor amplitudes on the same scale.

The derivative-to native structure factor amplitude scale factor was calculated with *SCALEIT* (using the Wilson option); this program applied Wilson's method to determine the absolute scale factor k (Wilson, 1942; Blundell and Johnson, 1976; Wolf *et al.*, 1991):

$$k_{\text{Wilson}} = \sqrt{\frac{\sum |F_p|^2}{\sum |F_{PH}|^2}}$$

Finally, instead of applying Wilson's method, the *SCALEIT* program determined an overall temperature factor B using a least-squares minimisation procedure.

5.2.4.2 Locating the Heavy atom Sites

The *FFT* program was used to calculate a difference Patterson map from the reflection data.

PEAKMAX was then used to look for possible heavy atom sites in the map: All positive peaks with a height above a given threshold were listed in grid and fractional coordinates.

The difference Patterson map was then plotted using *NPO*. After conversion to postscript format using the *PLTDEV* program, the Harker section(s) were then checked for peaks corresponding to those found in *PEAKMAX*.

5.2.4.3 Refinement of Heavy Atom Parameters

If any heavy sites were found, *MLPHARE* was used to refine the heavy atom parameters and error estimates, and then use these refined parameters to generate phase information.

MLPHARE refinement utilises the maximum likelihood principle. A basic treatment of this principle is that a joint probability density function is constructed which measures the likelihood that the particular set of measurements that were actually obtained, would have been obtained given any specified set of values for the unknown parameters. The optimal set of parameters is then those which maximise the likelihood of having made the actual set of measurements. With this method of refinement all possible values of the phase are considered, each value being weighted according to its probability of being correct. The weighted values are then minimised between the observed and calculated values of F_{PH} generated with this phase, with respect to the heavy atom parameters. The parameters are then up-dated and another cycle of refinement is carried out.

5.2.4.4 Locating and Refining possible Heavy Atom Sites in the H205Q D-HicDH 'Derivatives'

Only four of the data sets collected were considered to be complete enough (> ~85%) and to a sufficiently high resolution (> ~4.5Å) for attempts to be made at locating and refining possible heavy atom sites.

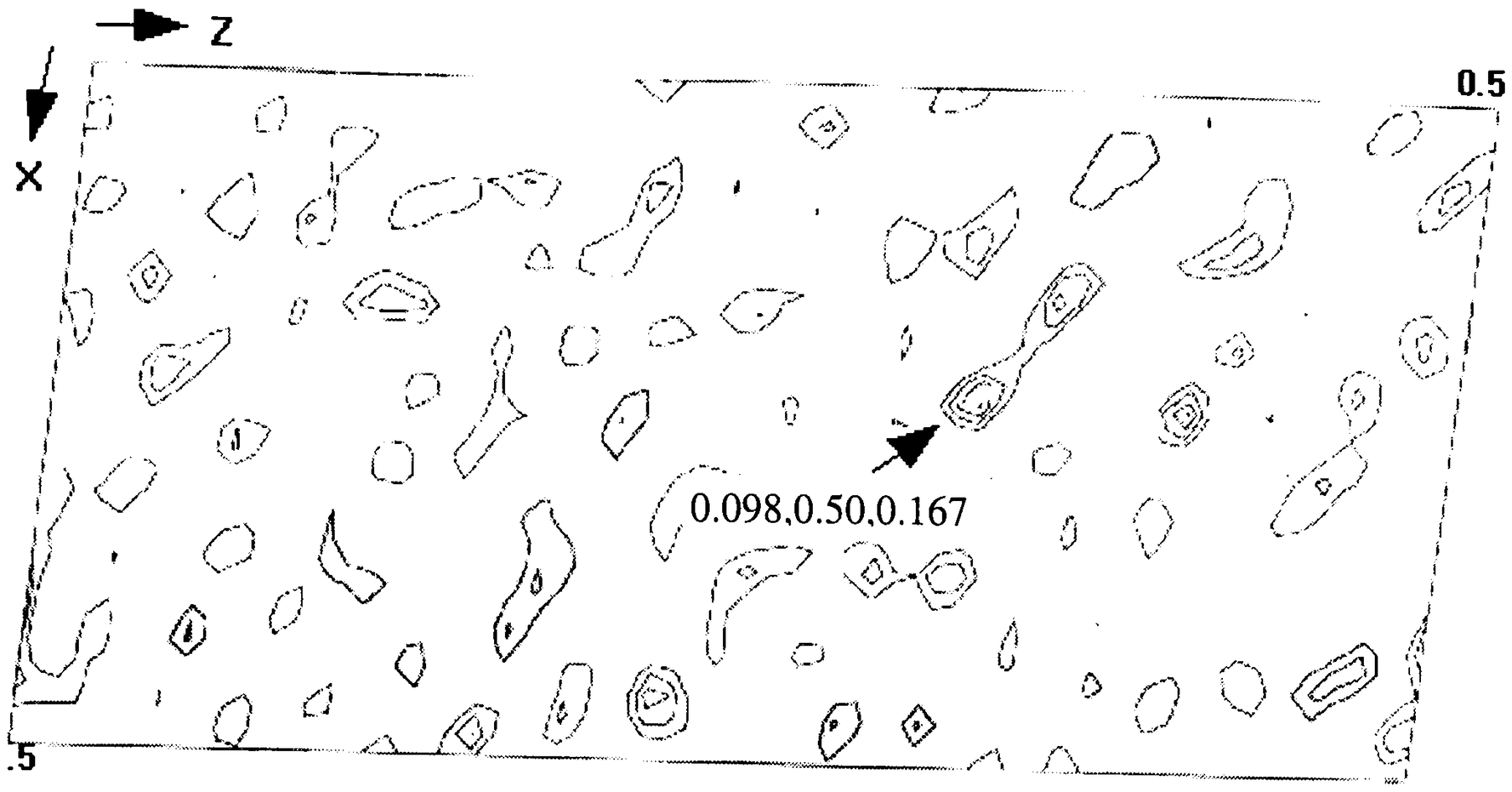
Derivative ⇒ number	2 (uranyl acetate)	3 (platinum potassium nitrate)	8 (platinum potassium nitrate)	17 (samarium chloride)
Mean R_{iso} (%)	21.3 (with native data to 3.6Å)	11.5 (with native data to 4.0Å)	23.2 (with native data to 4.0Å)	29.1 (with native data to 3.6Å)
First site				
x	0.099	No	No	0.098
y	0.50	peaks	peaks	0.50
z	0.213	visible	visible	0.167
Relative occupancy	0.273	-	-	0.905
R_{Cullis}				
centric	0.93	-	-	0.83
acentric	0.97	-	-	0.92
Phasing power				
centric	0.34	-	-	0.56
acentric	0.45	-	-	0.89

Table 5.16 Attempts at the location and refinement of possible heavy atom sites in the H205Q D-HicDH soaked crystals.

Peaks in the Harker sections, and *PEAKMAX*, were not found in the H205Q D-HicDH crystals soaked in platinum potassium nitrate.

The crystals soaked in uranyl acetate had too high an R_{Cullis} for centric reflections and a relatively low a phasing power for them to be considered useful in phasing.

The crystal which had been soaked samarium chloride appeared to have bound samarium with high occupancy (Figure 5.1).



Harker section v = ½

Figure 5.1 Harker section (v = ½) of the scaled native and samarium chloride structure factor amplitude difference Patterson map.

However, the R_{iso} was rather high (29.1%) suggesting that the nonisomophorphism was too great (Native $P2_1$ crystal; a=77.41, b=112.21, c=81.14, α=90°, β=96.81°, γ=90°. $SmCl_3$ soaked $P2_1$ crystal; a=76.73, b=113.89, c=81.14, α=90°, β=97.55°, γ=90°): Therefore the accurate location of any bound heavy atom would be very difficult. The problems were compounded by the fact that the data had a high R_{Cullis} for centric reflections and far too low a phasing power, and thus the crystal could not be used as a derivative.

5.2.5 Conclusions

SOLVE, an automated structure solution program, was also employed in searching for potential heavy atom derivatives; this program was also unable to find any suitable candidates.

Numerous attempts at preparing a heavy atom derivative had all been unsuccessful. This was due to the poor diffraction by the soaked crystals,

high mosaicity, nonisomorphism and / or low occupancy of the heavy atoms within the crystals.

5.3 Multiwavelength Anomalous Dispersion

5.3.1 Introduction

Selenomethionine derivatives can be used in the MAD phasing of a protein (as discussed in Chapter 2). *L.bulgaricus* D-HicDH contains 10 methionine residues (Figure 1.14) out of a total of 333 amino acid residues. After, comparing the sequence to known D-2-Hydroxyacid dehydrogenase structures it was estimated that nine methionine residues were probably ordered and would thus give a usable anomalous signal if substituted by selenomethionines; the remaining methionine residue is on the N-terminus of the polypeptide and therefore is probably disorded. Bearing in mind that some of the D-HicDH crystal forms contained multiple molecules in the asymmetric unit, the number of anomalous signals could be greater.

5.3.2 Preparation of selenomethionyl D-HicDH protein for phase determination

In order to prepare selenomethionyl (Se-Met) protein the cells expressing the protein of interest should be auxotrophic for methionine. One method of creating an auxotrophic strain is to inhibit the biosynthetic pathway for methionine. Alternatively, another approach is to tranform a pre-existing methionine autotrophic strain with the plasmid carrying the gene of interest (Doublié, 1997).

5.3.2.1 Methionine Biosynthetic Pathway Inhibition

5.3.2.1.1 Inhibiting Aspartokinases

High concentrations of isoleucine, lysine and threonine are known to block methionine biosynthesis in *E.coli* by inhibiting aspartokinases. Furthermore, phenylalanine and leucine act in synergy with lysine (Doublié, 1997; Voet and Voet, 1995; Van Duyne *et al.*, 1993). Selenomethionyl protein can thus be expressed by a nonauxotrophic prokaryotic strain if it is grown in a medium without methinone but with selenomethionine and large amounts of the amino acids known to inhibit methionine biosynthesis. This recently developed technique does not require the making of new *met*$^-$ expression vectors and so neither transformation nor transduction of the plasmid is needed. This procedure has been successfully employed in the determination of several structures (e.g. Doublié *et al.*, 1996; Benson *et al.*, 1995; Van Duyne *et al.*, 1993).

5.3.2.1.2 Bacterial Growth in Selenomethionine Containing Media

It was found that bacterial growth in selenomethionine could be unpredictable unless certain precautions were taken. One problem was that cells that were permitted to reach stationary phase often died, and could not then be used to innoculate larger cultures. A possible explanation for this is that selenomethionine is incorporated almost entirely into the protein while the cells are in the exponential growth phase (Huber *et al.*,1967).. However, during stationary phase, selenium in toxic forms may be released on the breakdown of selenomethionine.

Another problem is that cells grow more slowly in minimal media containing selenomethionine, reaching stationary phase at a lower cell density, and having a longer lag time (Huber *et al.*, 1967; Cowie and Cohen,

1957). Slow growth can also lead to the potential problem of oxidation of the selenomethionine (Doublié *et al.*, 1996); oxidation can lead to a loss of birefrigence and diffraction of the selenomethionyl protein crystals (Hendrickson *et al.*, 1990).

It has been suggested that cells growing in selenomethionine need a small amount of rich medium, for example Luria broth, to get out of stationary phase (Yang *et al.*, 1990). In order to drastically reduce the concentration of the methionine from the starter culture, the cells are then used to innoculate a larger volume of selenomethionine containing, methionine-free, minimal media, which in turn is used to innoculate a much larger volume of minimal media supplemented with selenomethionine. The final dilution of the methionine in the starter culture has to be as great as possible because the cells preferentially incorporate methionine rather than selenomethionine (Huber *et al.*, 1967).

5.3.2.1.3 Preliminary Investigations into AR58p[GIN113/6]Bacterial Growth on Selenomethionyl Rich Media

(i) In order to determine the best conditions for initial cell growth, a glycerol stock of the clone harbouring the H205Q mutant D-HicDH was streaked onto two agar plates: One of these was a rich 1.6% LB agar plate and the other a minimal media agar plate supplemented with seleno-L-methionine (Sigma; S3132):-

Preparing Rich Luria Bertani Agar (100ml)
(From Dept. of Microbiology and Pathology, University of Bristol)

Agar	1.65 g
Bacto-tryptone	1 g
Bacto-yeast extract	0.5 g
Ampicillin	0.1 mg

(pH to 7.2-7.5 and autoclave for 15mins)

Preparing Minimal Agar (100ml)

Agar	1.65 g
Glucose	0.2 g
$MgSO_4$	20 mg
Thiamine	0.5 mg
1 x M9minimal salts	(1.28g $Na_2HPO_4 \cdot 7H_2O$, 0.g KH_2PO_4, 100mg NH_4Cl, 50mg NaCl)
Ampicillin	0.1 mg
seleno-L-methionine	6 mg

(pH to 7.0-7.2 and autoclave for 15mins)

The plates were incubated at 37°C. Overnight many single bacterial colonies had grown on the LB agar plate. However, even after 3 days, there were still no visible colonies on the minimal agar plate.

(ii) Single colonies from the rich LB agar plate were used to innoculate 10ml of rich Luria broth and 10ml M9 minimal media supplemented with seleno-L-methionine (Sambrook *et al.*, 1989). The same procedure was repeated using colonies from the minimal agar plate (as there were no visible colonies, samples were transferred from regions of the tray where indentations had been made whilst streaking the glycerol stock).

Preparing Rich Luria Media (100ml)
(From Dept. of Microbiology and Pathology, University of Bristol)

Bacto-tryptone	1 g
Bacto-yeast extract	0.5 g
Ampicillin	0.1 mg

(pH to 7.2-7.5 and autoclave for 15mins)

Preparing M9 Minimal Media (100ml)

Glucose	0.2 g
1 x M9minimal salts	(1.28g $Na_2HPO_4 \cdot 7H_2O$, 0.g KH_2PO_4, 100mg NH_4Cl, 50mg NaCl)
Ampicillin	0.1 mg
seleno-L-methionine	6 mg

(pH to 7.0-7.2 and autoclave for 15mins)

All four samples were placed overnight in a shaking incubator set at 37°C.

The cell density of each sample was then measured:-

Sample	OD_{600nm}
RM (Colony from rich agar innoculating 10ml M9 minimal media)	0.11
MM (Colony from minimal agar innoculating 10ml M9 minimal media)	0.001
MR (Colony from minimal agar innoculating 10ml M9 rich L media)	0.01
RR (Colony from rich agar innoculating 10ml rich L media)	1.81

Table 5.17 AR58p[GIN113/6] cell densities in various 'start' media

These results show that the AR58p[GIN113/6] bacterial strain expressing D-HicDH is unable to grow on M9 minimal media agar containing Se-Met.

(iii) The 10ml cultures that grew, namely samples RM and RR, were then used to innoculate 500ml M9 minimal media supplemented with Se-Met: The samples were called RMM and RRM, respectively. The cell density of both samples were then measured at various time intervals (Figure 5.2)

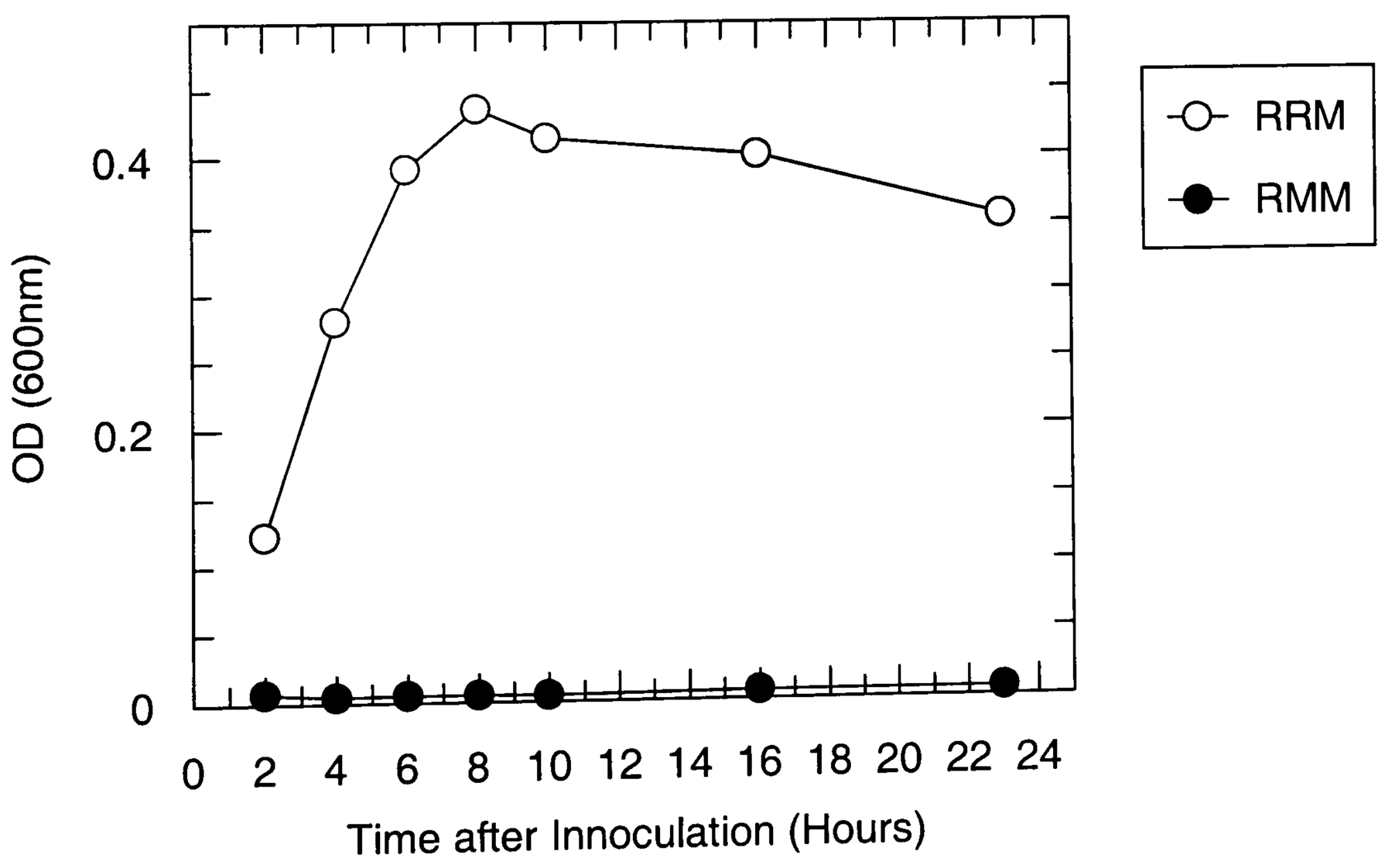


Figure 5.2 Typical growth curve for AR58p[GIN113/6] bacterial cells in selenomethionine rich media.

The results seem to suggest that the cells, growing in selenomethionine, do indeed require a small amount of rich medium to get out of lag phase (as proposed by Yang *et al.*, 1990).

(iv) The whole procedure was repeated a further two times in order to obtain 'average' cell growth curves.

It was concluded that, in order to obtain the optimum Se-Met protein expression, colonies grown on LB agar should be used to innoculate 10ml L broth; which in turn should be to innoculate 500ml M9 minimal media supplemented with Se-Met. Protein expression should then be induced after the cell had reached an OD_{600nm} = 0.3-0.35; this was calculated to be the stage at which the cells were in the mid-log phase of growth. The cells should then be harvested at an OD_{600nm} = 0.4-0.45 (just before the onset of stationary phase).

5.3.2.1.4 Se-Met D-HicDH Purification Protocol using the AR58p[GIN113/6] Expression System

Using aseptic techniques, a glycerol stock of AR58p[GIN113/6] was streaked out onto a 1.6% LB agar plate and grown overnight at 37°C. A single bacterial colony was toothpicked from the plate and used to inoculate 10ml L Broth, this was again grown at 37°C in a shaking incubator overnight. All of the culture was then transferred to a sterile 2l flask containing 500ml M9 minimal media, which was then grown at at 37°C until the cells had reached an OD_{600nm} of 0.33.

When the culture had reached the mid-log phase of growth, an amino acid 'cocktail' consisting of lysine, phenylalanine and threonine at 100 mg/litre; isoleucine, leucine and valine at 50 mg/litre, was added to the culture to inhibit methionine biosynthesis: Also, seleno-L-methionine at 60 mg/litre

was added (Van Duyne *et al.*, 1993). 15 minutes after the addition of amino acids, protein expression was induced (by heat shocking the culture at 42°C for 3 hours and thus derepressing the lambda repressor). The cells were allowed to grow until an OD_{600nm} of 0.42 had been reached.

The cells were then harvested and the D-HicDH protein purified in a very similar way to that for the native enzyme. However, to avoid oxidation of the seleneomethionine residues, all the buffers used were degased, and fresh 5mM dithiothreitol (DTT) and 0.2mM ethylenediaminetetra-acetic acid (EDTA) added to them. EDTA is a chelator that forms a complex with any metal ions in solution that might aid in oxidation (Stoll and Blanchard, 1990).

Protein samples were taken at various stages during the purification procedure. After protein concentration, the subsequent SDS-PAGE expression gels were very similar to that of the native (Figure 4.3), however, the protein yield had been dramatically reduced. Only about 1mg of purified D-HicDH was obtained from a 2 litre culture (Table 5.18), whereas during a 2 litre native D-HicDH protein preparation 100-150mg of enzyme was obtained.

Stage	Volume (ml)	Total protein (mg)	Total activity (U)	Specific activity (U/mg)	Yield (%)	Purification factor -fold
Crude cell extract	31	59.3	298	5.0	100	1.0
After acid dialysis	26	33.4	291	8.7	97.7	1.74
After cationic exchange (S - column)	58	24.3	221	9.1	74.2	1.8
After anionic exchange (Q - column)	64	1.2	24.6	20.5	8.3	4.1

Table 5.18 D-HicDH selenomethionyl protein purification statistics from a 2 litre AR58p[GIN113/6] culture.

Compared with the native protein preparation, there was an enormous reduction in the total amount of protein isolated from the crude cell extract during selenomethionyl D-HicDH purification. This could be partly attributed to the differences in the cells densities reached by the cultures of the two preparations. However, there was also a substantial drop in the yield of D-HicDH obtained during the penultimate step of the purification process : About 20% of the protein was lost after S-sepharose cationic exchange, whereas, during native protein purification only about 10% of the enzyme had been lost after the sample had passed through the S-sepharose column. Furthermore, 65% of the protein was lost after the Q-sepharose step in the procedure, whereas during the native purification preparation only 10% of D-HicDH was lost after anionic exchange. Therefore, there was insufficient protein for crystallisation trials and mass spectrometric analysis, and more enzyme had to be purified.

5.3.2.1.5 Using JM105p[KK183] Cells and Catabolite Repression to Reduce Pre-Induction D-HicDH Expression

Smith (1991) suggested that in order to obtain a satisfactory anomalous signal at least 98% of the methionine residues must be substituted for selenomethionine in the derivative crystal. It was soon realised that the final dilution of the initial methionine from the rich medium (~1:50) was probably not great enough to obtain suitable crystals for MAD phasing. Furthermore, it was also known that the pGIN plasmid (whose inserted gene promoter is controlled by a lambda repressor) expressed low levels of D-HicDH protein even when the cells were uninduced. As methionine is preferentially incorporated into proteins and seleno-L-methionine is added to the culture only a few minutes before induction, this low level of pre-induction protein expression would inevitably lead to the translation of some native enzyme.

In order to carry out mutagenic studies on the H205Q D-HicDH protein, the gene encoding the enzyme had previously been subcloned into a pKK expression vector (pKK 183) and then transformed into an *E.coli* JM105 host strain (K.Moreton, unpublished results). As well as conferring ampillicin resistance, the pKK plasmid also contains a *tac* promotor for over-expression of recombinant proteins. The *tac* promotor is very similar to the *lac* promotor and is inducible with isopropyl-β-D-galactopyranoside (IPTG), an analogue of lactose (Turner *et. al.*, 1997).

The JM105 bacterial strain lacks the *lacI* gene (*i.e.* it is a lac I^q strain) that encodes the *lac* repressor which is a potent inhibitor of trancription (Watson *et al.*, 1992). In the absence of IPTG, the *lac* repressor would bind to the *tac* operator, just upstream of the promoter, and repress transcription from the *tac* promoter. Upon induction, an IPTG-repressor complex, which is unable

bind to the operator site, is formed allowing expression of the H205Q D-HicDH gene product at a very low level.

Previous protein expression experiments had shown that very little D-HicDH was produced prior to induction in JM105p[KK183] cells (Figure 5.3; K.Moreton, unpublished results):-

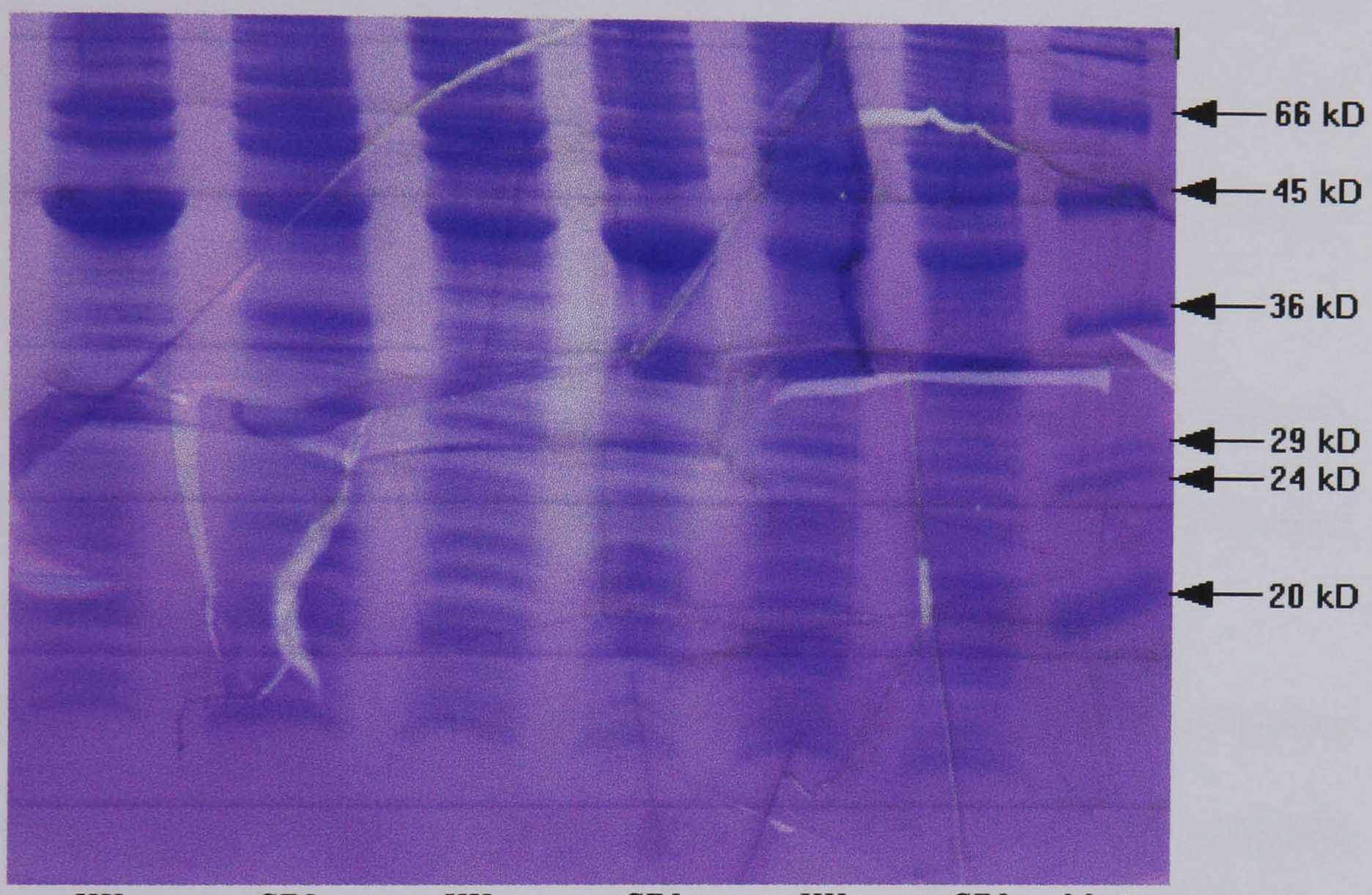


Figure 5.3 SDS- PAGE analysis of D-HicDH expression. This 12.5% acrylamide gel shows protein expression 3 hours after induction. The AR58p[GIN113/6] and JM105p[KK183] expression systems were either induced with 1mM IPTG, or heat-shocked at 42°C for 3 hours, or the cells remained uninduced.

The JM105p[KK183] expression system seemed ideal for selenomethionyl protein production. It showed both a lower level of pre-induction protein expression when compared with AR58p[GIN113/6], and the *tac* promoter could be regulated by catabolite repression.

Transcription from the P_{tac} promoter is not only dependent on the presence of IPTG / lactose but also on the concentration of glucose in the culture (Makman and Sutherland, 1965; Watson *et al.*, 1992, Turner et al., 1997).

The decreased rate of transcription from operons in the presence of glucose, termed catabolite repression, is mediated by an allosteric protein that binds to cyclic adenosine 3′,5′-monophosphate (cAMP) and is thus called cAMP receptor protein (CRP). In the absence of glucose the cAMP concentration in *E.coli* cells rise (Voet and Voet, 1995). CRP binds to cAMP to form a cAMP:CRP complex which is able to bind to a specific sequence near the *tac* promoter which is just upstream from the site for T7 RNA polymerase. The protein complex binding induces a 90° bend in the DNA, which is thought to enhance RNA polymerase binding to the promoter and thus enhance transcription.

The cAMP levels in *E.coli* cells growing in in the presence of glucose is low, and CRP exists in an inactive form unable to bind DNA. Therefore transciption is initiated inefficiently in the presence of glucose.

Protein expression levels in JM105p[KK183] in the presence of excess glucose, namely 1% glucose (Makman and Sutherland, 1965), were analysed and compared with expression of E.coli cells grow in cultures containing 'normal' glucose concentrations (*i.e.* 0.2% glucose) (Figure 5.4).

Figure 5.4 SDS- PAGE analysis of D-HicDH expression in JM105p[KK183] in the presence 0.2% or 1% glucose. Samples were either taken just prior to induction or 4 hours after.

In the presence of excess glucose there was very little pre-induction D-HicDH expression. It was also observed that there was still a substantial amount of expression of protein after induction with IPTG even when the culture contained 1% glucose.

5.3.2.1.6 Se-Met D-HicDH Purification Protocol using the JM105p[KK183] Expression System

Initially, preliminary experiments were carried out to determine the typical growth curve for the JM105p[KK183] cells in M9 minimal media containing seleno-L-methionine and 1% glucose.

After the cell growth conditions had been established, a procedure was devised that would this time include a much larger dilution of the initial methionine in the starter media.

A single JM105p[KK183] bacterial colony grown on a 1.6% LB agar plate was used to innoculate 5ml L broth containing a further 1% glucose, this was grown overnight at 37°C. 250µl of this culture was then used to innoculate 5ml M9 media. This minimal media, along with all the subsequent media in this protocol, contained 60mg/l seleno-L-methionine and 1% glucose.

After the culture had reached the desired cell turbidity (OD_{600nm} of 0.3-0.35) 2.5ml was used to innoculate 10ml M9. All of this culture was later used to innoculate 25ml M9. And in turn, when the correct cell density had been reached, this culture was poured into a 2 litre flask containing 100ml minimal media. This was then used to innoculate 500ml M9 minimal media. Each step in this process could take up to two days, and so therefore the ampillicin in the media was replenished every 24 hours. The final dilution of the initial methionine was about 1:10,000.

The 625ml M9 minimal culture was allowed to grow until an OD_{600nm} of 0.3-0.35, then the amino acids that inhibit methionine biosynthesis were added (at the previous concentrations mentioned above). 15 minutes later, protein expression was induced with the addition of 1mM IPTG. The cells were grown until they had reached an OD_{600nm} of 0.42-0.46, and then harvested. D-HicDH was then purified as before. Typically, 12-17mg of purified D-HicDH was obtained from a 2 litre culture.

Stage	Volume (ml)	Total protein (mg)	Total activity (U)	Specific activity (U/mg)	Yield (%)	Purification factor -fold
Crude cell extract	42	96	556	5.8	100	1.0
After acid dialysis	58	54	491	9.1	88.3	1.6
After cationic exchange (S - column)	62	42	372	8.9	66.9	1.5
After anionic exchange (Q - column)	98	16.4	348	21.2	62.6	3.7

Table 5.19 Typical D-HicDH selenomethionyl protein purification statistics from a 2 litre JM105p[KK183] culture.

During the purification preparations using JM105p[KK183] cells there was only about a 5% loss in yield after the protein had been eluted from the Q-sepharose column. However, there was still a substantial loss in protein after cationic exchange.

After sonication, there was a greater amount of total protein and D-HicDH activity obtained from the crude cell extract in these preparations when compared with the selenomethionyl protein purification from AR58p[GIN113/6] cells. However, when induced, the AR58p[GIN113/6] cells tended to have a greater level of protein expression than the JM105p[KK183] cells in 1% glucose (*cf.* Figures 5.3 and 5.4). These two observations seem to suggest that the *E.coli* cells prefer a gradual drop in the methionine concentration rather than a sudden change to seleno-L-methionine.

The biochemical company 'Novagen' had stated that the use of a protein extraction reagent, such as 'BugBuster', would liberate up to 8 times as much soluble protein from *E.coli* cells compared with conventional sonication. The detergent mixture works by disrupting the cell walls without

denaturing soluble protein. However, when BugBuster was employed there was slightly less protein in the soluble crude cell extract.

5.3.3 Mass Spectrometry

In order to ascertain the degree of selenomethionyl incorportation, the protein samples were analysed using mass spectrometry (MS). Two distinct types of MS were employed.

5.3.3.1 Electrospray Ionization Mass Spectrometry

In the electrospray process intact protein molecules in ionized form are produced from a dilute solution of the protein sample by creating a spray of highly charged droplets in the presence of a strong magnetic field. The protein solution is usually in a water-organic solvent with a trace of trifluoroacetic acid added to promote ionization.

During electrospray ionisation mass spectrometric (ESI-MS) analysis, the protein solution is sprayed from a region of strong electric field. The ionized droplets are then attracted to an inlet pipe which is held at an opposite potential. During their journey through a vacuum system, solvent contained in the droplets is caused to evaporate by applying a directed flow of dry gas.

The ionized protein molecules then undergo collisions with neutral gas molecules in the entrance of the vacuum system. The resulting protein molecules carry an excess positive charge related to their charge state in solution (Fenn *et al.*, 1989). The molecules are then directed into a quarupole filter (Chait and Kent, 1992) for mass / charge (m/z) ratio analysis. An algorithm is then used to convert the ion peaks into a single peak corresponding to the neutral molecule.

Only picomole amounts of protein are consumed using ESI-MS, so the sample should be at a concentration of 1-300 pmole/microliter. Mass

accuracies of ± 1 Dalton in 1,000 Daltons are commonly observed (Chait and Kent, 1992).

5.3.3.2 Matrix-Assisted Laser Desorption Ionization Mass Spectrometry

The protein sample is prepared for matrix-assisted laser desorption mass spectrometric (MALDI-MS) analysis by adding the solution to an aqueous solution containing a large excess of concentrated matrix-forming material such as 3,5-dimethyoxy-4-hydroxycinnamic acid (Chait and Kent, 1992). A small volume of this mixture is dried on a probe which is then inserted into the mass spectrometer.

The sample now consists of a layer of matrix microcrystals containing isolated protein molecules. This is bombarded with short pulses of ultra-violet radiation produced by a laser. The laser pulse causes a fraction of the matrix and trapped protein molecules to go into a gas phase and also ionizes the protein. A static field is then imposed on the ions. This is generated by application of a high potential with respect to a closely spaced electrode. The ions are accelerated through a hole in the electrode into a electric field free flight tube (about 0.5 to 2 metres long).

Because the ions are accelerated through a fixed potential difference, the velocity of the ions is simply proportional to $(m/z)^{-½}$, and thus simple time of flight (TOF) analysis can be used to calculate the masses of the protein ions.

MALDI-MS requires only femtomole amounts of the protein and usually a sample concentration of 250 fmole - 500 pmole is used.

MALDI-MS is a very sensitive and the TOF analyser has to be calibrated with a protein of known mass with a similar molecular weight to that of the protein sample. Assuming the mass spectrometer is properly calibrated,

MALDI has a mass accuracy of ± 1 Dalton in 10,000 Daltons (Chait and Kent, 1992).

5.3.3.3 Mass Spectrometric Analysis of Selenomethioyl D-HicDH Protein Produced by Inhibiting Methionine Biosynthesis

Before protein samples can be analysed using mass spectrometry involatile salts and solvents found in the sample buffers must be removed. Therefore 0.5ml samples of the native and putative selenomethionyl H205Q D-HicDH protein were dialysed against 5 litres of milli Q water overnight. However, it was clear from the subsequent electrospray ionization mass spectrometric analyses performed by John Crosby (Department of Chemistry, Bristol University) that there was still a substantial amount of salt still left in the samples (Figure 5.5).

Other methods for removing salt include reversed-phase HPLC and size-exclusion chromatography. However, these methods are not suitable for desalting small quantities of protein (1-100 pmol). Firstly, these procedures (along with dialysis) dilute the protein samples which then have to be concentrated before MS analysis. And secondly, these desalting methods require multiple handling steps.

The protein samples were concentrated and desalted using the approach devised by Winston and Fitzgerald (1998): Approximately 1-5mg of reversed-phased C4 chromatographic media (which binds protein) was suspended in 100µl of methanol. This mixture was added to the dialysed protein solution, and the sample vortexed for 5 minutes. The samples were then spun for 3 minutes at 14,000rpm in order to pellet the C4 media. 10µl of methanol was then added to the surface of the sample to complete precipitation.

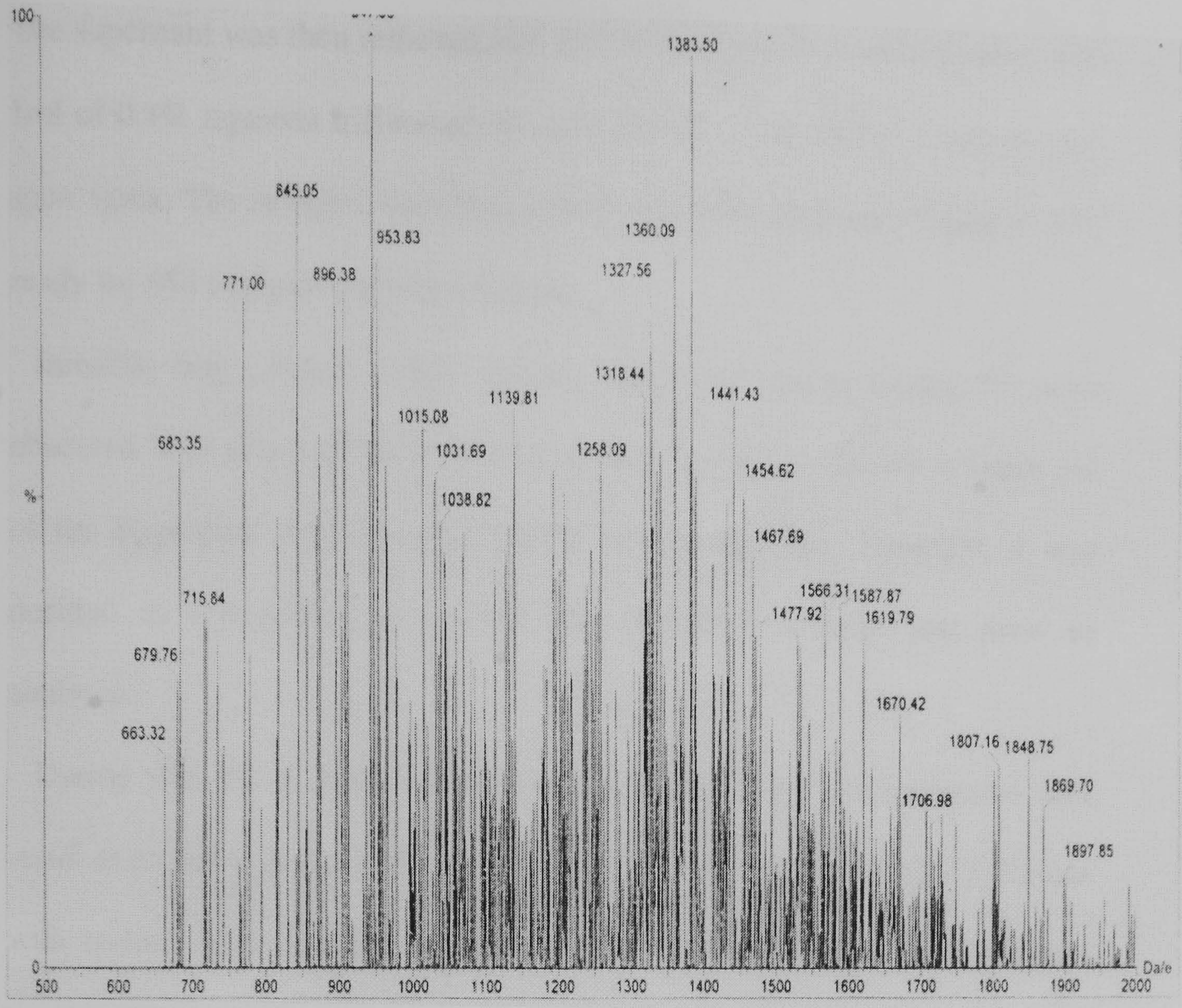


Figure 5.5 ESI mass spectrum of a protein sample just dialysed against water.

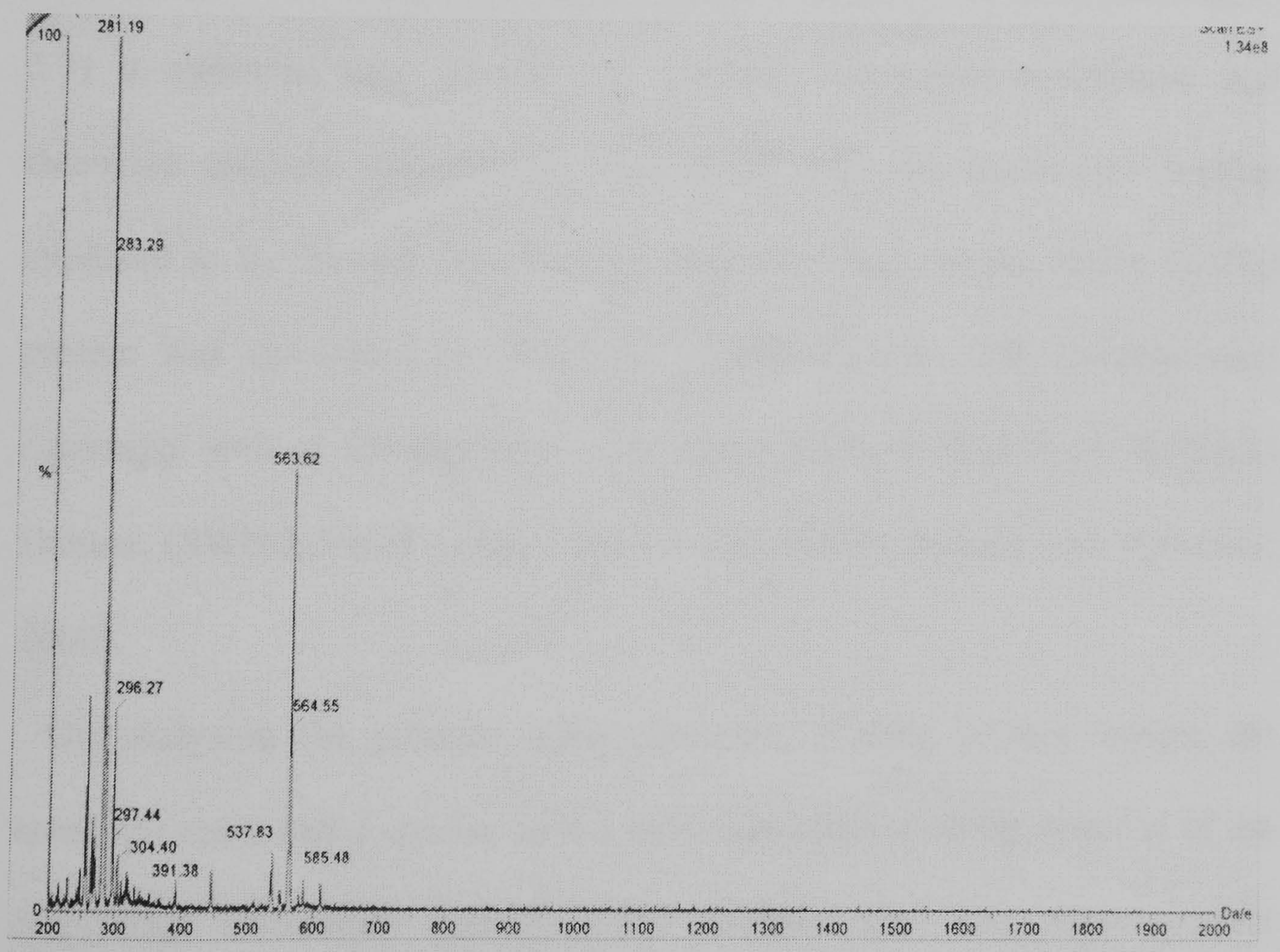


Figure 5.6 ESI mass spectrum of a protein sample stored in TFA overnight. It appears plasticiser has leached into the sample from the containing Eppendorf tube.

The supernant was then removed and the remaining pellet washed twice with 1ml of 0.1% aqueous trifluoroacetic acid (TFA), vortexed for 1 minute, and spun again. The samples were then stored in a 0.5ml Eppendorf tube at 4°C, ready for MS analysis the following day.

Initially, only protein sample spectra similar to that in Figure 5.6 were observed. It is suspected that the TFA solvent caused plasticiser to leach out of the Eppendorf tubes containing the protein samples. Therefore it was decided to concentrate and desalt the D-HicDH samples just prior to analyses.

During ESI-MS it is not uncommon for some of the termini amino acid residues to get knocked off whilst the protein molecules undergo collisions with residual neutral gas molecules in the entrance region of the vacuum system (Chait and Kent, 1992).

When the native H205Q D-HicDH protein sample was analysed (Figure 5.7) it appeared that 'species A' had the N-terminal methionine and threonine residues removed. In 'species B' only the methionine residue appeared to be missing. The average molecular mass of the native H205Q protein was calculated to be around 36,938 Daltons. This corresponded extremely well to the theoretical mass for H205Q D-HicDH of 36,939.34 Daltons (SWISS-PROT protein sequence database; Bairoch and Apweiler, 2000).

On analysing the putative selenomethionyl H205Q protein sample the spectrum again had a species with a mass consistent with the removal of the N-terminal methionine and threonine residues from the native protein molecule (Figure 5.8). However, there was also a peak in the spectrum that appeared to correspond to H205Q again having only its N-terminal

methionine removed but this time with the remaining 9 methionine residues all

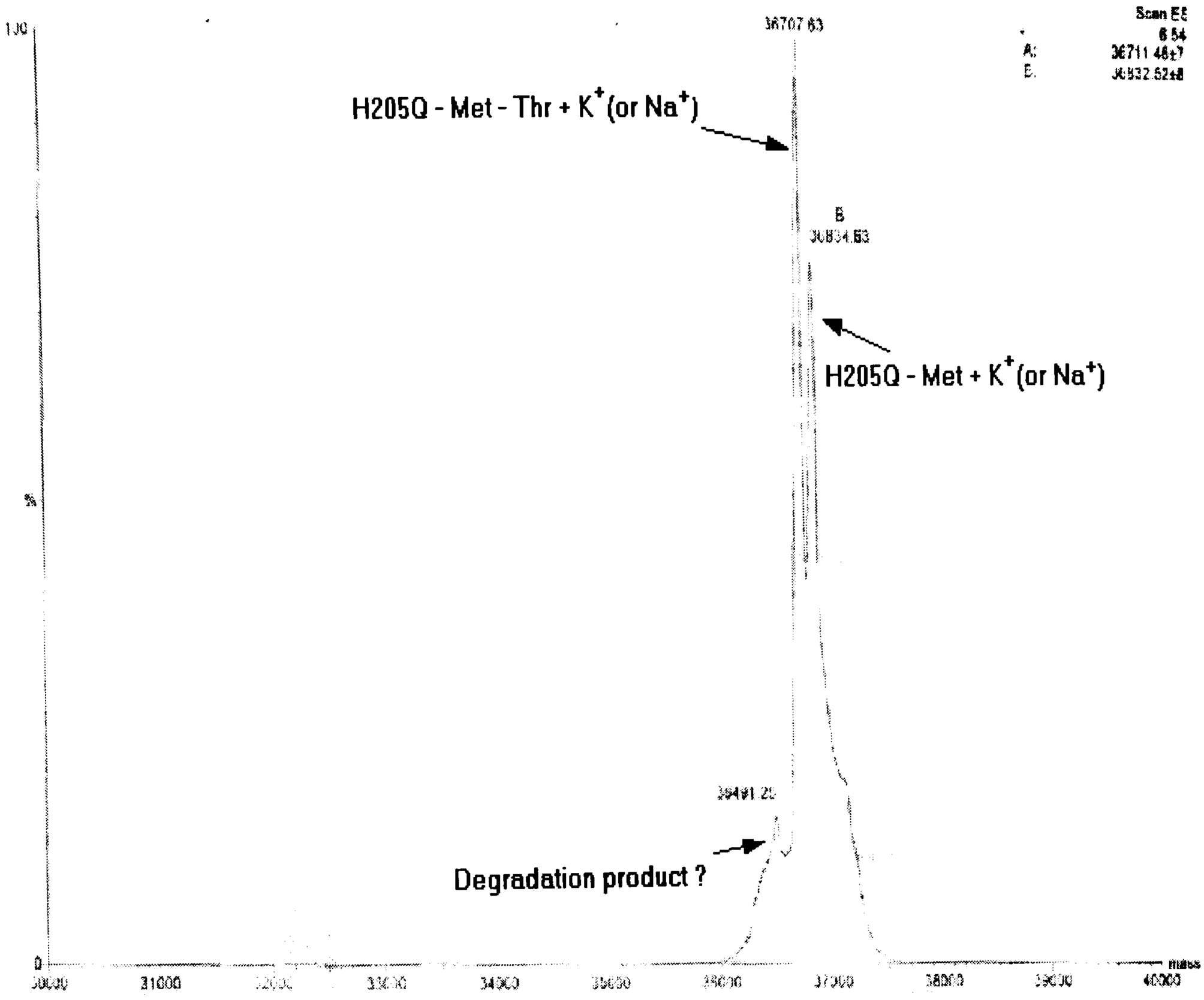


Figure 5.7 ESI mass spectrum of the native H205Q D-HicDH protein sample

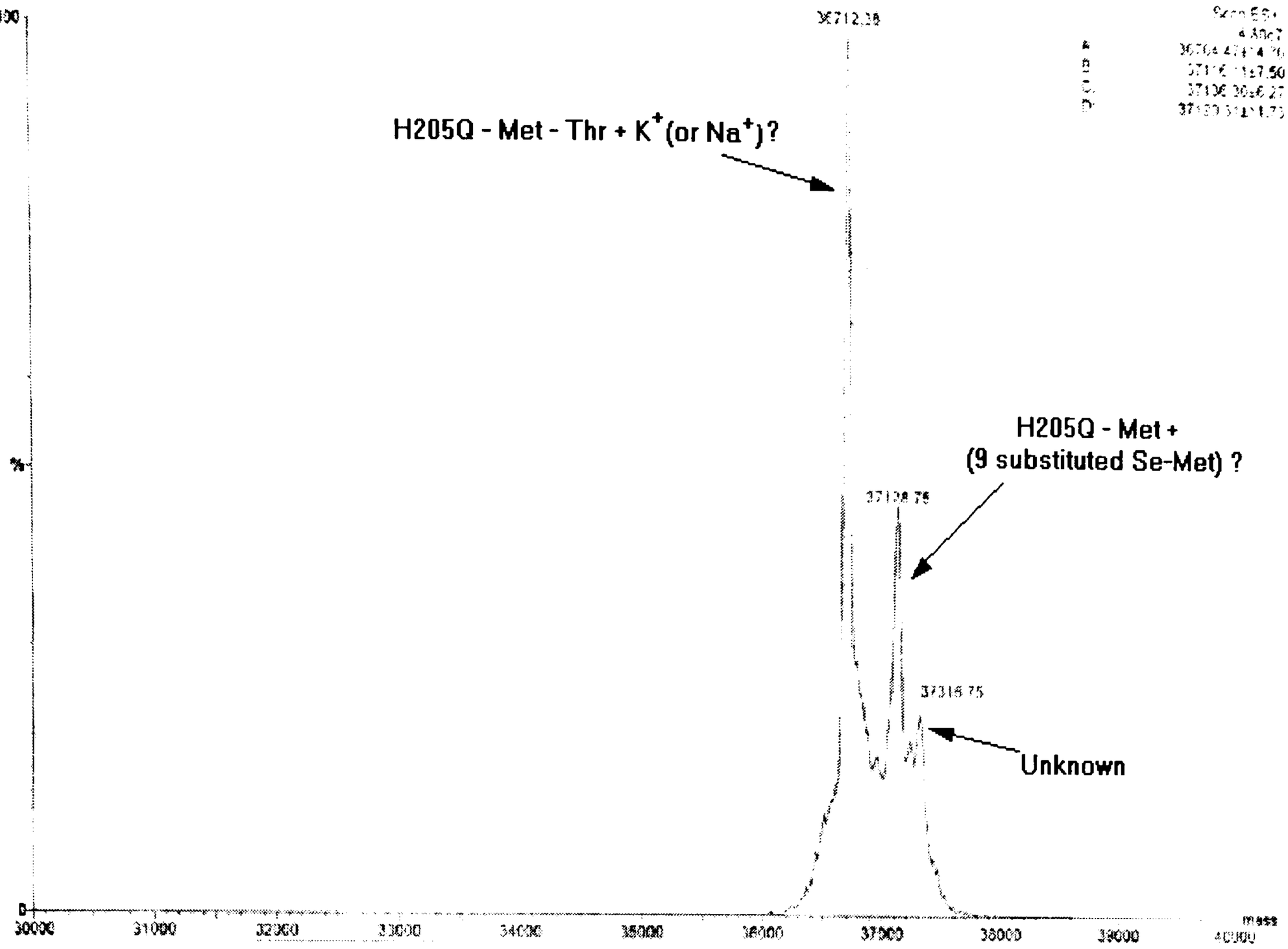


Figure 5.8 ESI mass spectrum of the putative selenomethionyl H205Q D-HicDH protein sample produced by inhibiting methionine biosynthesis.

being substituted for selenomethionine residues. The mass spectrum seems to suggest that there was an all-or-nothing substitution of methionine residues for seleno-L-methionine.

During expression of the 'selenomethionyl' protein, the inintial methionine in the cultures was diluted by about 1:10,000. Therefore, it appears that the methionine biosynthetic pathway was not completely inhibited by the excess concentrations of specific amino acids.

5.3.4 Selenomethionyl Protein Expression using a Methionine Auxotroph

Attempts at fully incorporating selenomethionine into D-HicDH by inhibiting methionine biosynthesis proved to be only partially successful. In order to ensure 100% Se-Met incorporation into the protein, the pKK 183 plasmid needed to be transformed into a preexisting *E.coli* methionine auxotrophic strain.

5.3.4.1 The DL41 and DL41(DE3) Methionine Auxotrophic *E.coli* cell strains

DL41 and DL41(DE3) are *E.coli* methionine auxotrophic cells strains. Glycerol stocks of both strains were a gift from Mark Banfield (School of Biological Sciences, Auckland University, N.Z.).

The DL41 and DL41(DE3) strains are derived from *E.coli* K12, and were constructed by LeMaster by introducing a lesion into the *metA* gene coding for homoserine O-succinylase (Hendrickson *et al.*,1990). There is therefore a block in the biosynthetic pathway for methionine between homoserine and O-succinylhomoserine (Figure 5.9).

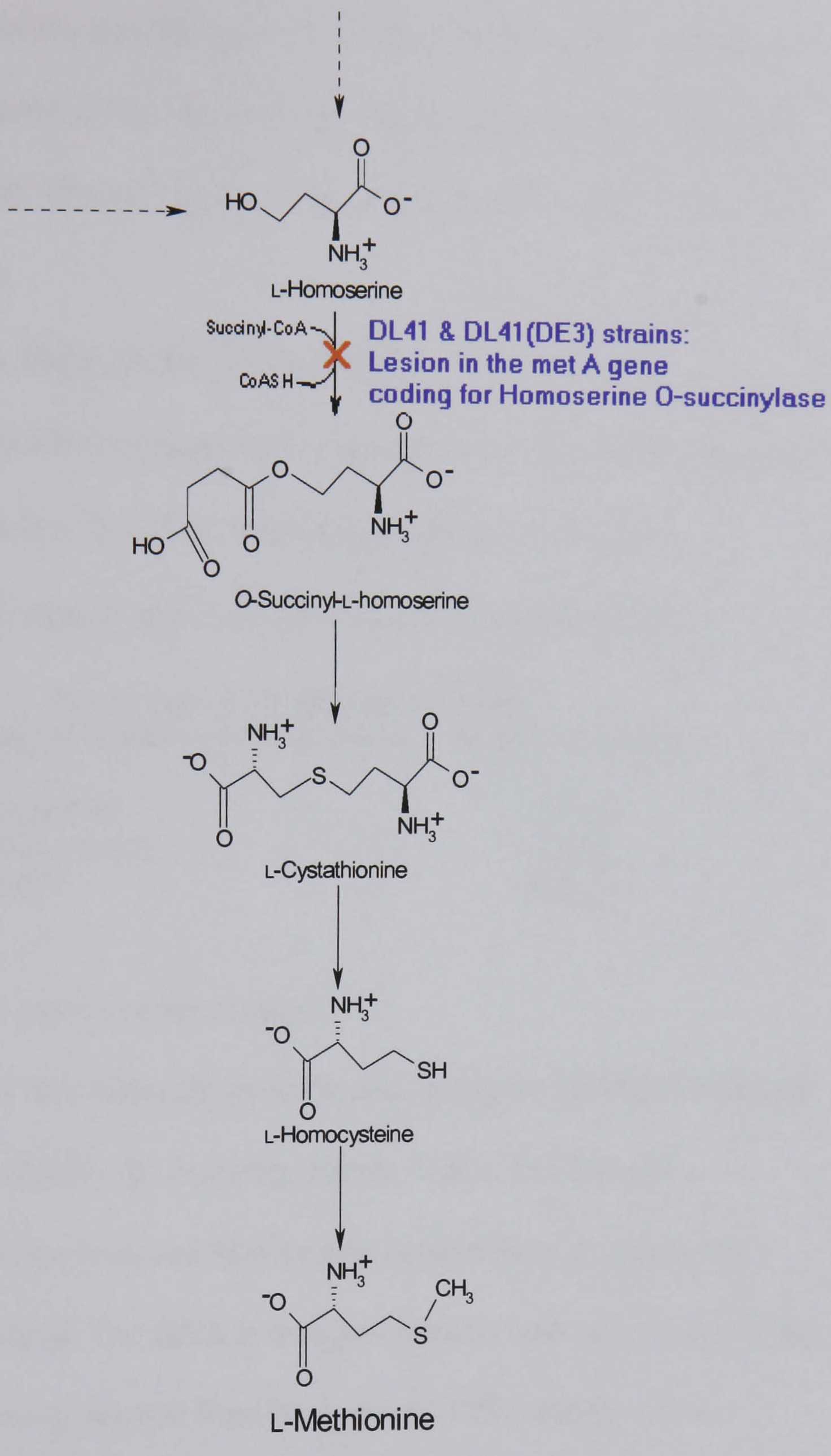


Figure 5.9 Lesion in the *metA* gene corresponding to a block in the methionine pathway in DL41 and DL41(DE3) cells.

The designation (DE3) indicates that the DL41(DE3) is a lysogen of λDE3, and therefore carries a chromosomal copy of the T7 RNA polymerase gene under the control of the *lacUV5* promoter. Therefore expression of this gene can also be induced by IPTG, thus making transcription of the H205Q gene even more efficient. Thus it was decided to transform the pKK 183 into the DL41(DE3) strain.

5.3.4.2 pKK 183 Plasmid DNA Extraction

A single JM105p[KK183] bacterial colony grown on a 1.6% LB agar plate was used to innoculate 5ml 2×YT medium (Sambrook *et al.*, 1989) containing 1μg/ml ampillicin and this was grown overnight at 37°C.

Preparing 2×YT Medium (100ml)
(From Dept. of Microbiology and Pathology, University of Bristol)

Bacto-tryptone	1.6 g
Bacto-yeast extract	1.0 g
NaCl	0.5 g

(pH to 7.2-7.5 and autoclave for 15mins)

Plasmid DNA was then extracted from the cells using the QIAGEN Plasmid Miniprep Kit (QIAGEN Ltd., Dorking, Surrey, U.K.). The kit uses a combination of alkline lysis and SDS protein denaturation to release the DNA from the bacteria. The DNA is then precipitated with isopropanol. The plasmid DNA is then separated from the genomic DNA using an anion exchange resin and differential salt washes.

The plasmid DNA extracted was then quantified on an 0.8% agrose gel, stained with ethhidium bromide, by comparison of band intensities with λHindIII markers (Turner *et al.*, 1997). It was found that the plasmid DNA had a concentration of about 50ng/μl and this was subsequently stored at -20°C.

5.3.4.3 Preparation of Competent DL41(DE3) Cells

(1) In order to make the cells competent a glycerol stock of DL41(DE3) was streaked onto a minimal agar plate supplemented with 80mg/l L-methionine. The plate was then incubated for 24 hours at 37°C; the colonies that grew were very small.

(2) A single colony was then used to innoculate 5ml 2×YT medium. (No antibiotics were added to the 2×YT as the DL41(DE3) cell strain does not carry an antibiotic resistance gene.)

(3) 0.5ml of the overnight culture was added to 50ml 2×YT (in a sterile 250ml flask). This was then grown at 37°C until the cells had reached an $O.D._{550nm}$ of about 0.3.

(4) The culture was then poured into pre-cooled sterile oakridge tubes, and the cells harvested by centrifugation at 4°C for 5 minutes at 4,000rpm.

(5) The cell pellets were each gently resuspended in pre-cooled 25ml $CaCl_2$ solution.

Preparing $CaCl_2$ Solution
100mM $CaCl_2$
5mM $MgCl_2$
5mM Tris

(pH to 7.6 and autoclave for 15mins)

From here onwards, the cell solutions were always kept cold.

(6) The cell suspension was then incubated on ice for 30 minutes, occassionally gently mixing the solution.

(7) The cells were spun down at 4°C for 5 minutes at 4,000rpm. And the resulting cell pellets each resuspended in pre-cooled 1ml $CaCl_2$ solution.

(8) The cells were then left on ice for 2 hours; during this time the competence of the cells increased.

5.3.4.5 Transformation of the pKK 183 plasmid into the Competent DL41(DE3) Cell Strain

(1) 200µl of the cell suspension was mixed with 100ng of the thawed plasmid DNA solution (*i.e.* 2µl). This was then incubated on ice for 30 minutes.

(2) The cells were then heat shocked by putting them in a water bath set at 42°C for 90 seconds.

(3) 250µl of pre-warmed 2×YT medium (at 42°C) was then added to the cells and incubated at 37°C for 1 hour.

(4) 350µl of the transformed cells were onto 1.6% LB agar plate (containing ampillicin) and grown overnight at 37°C.

5.3.4.6 Preparation of DL41(DE3)p[KK183] Glycerol Stocks

(1) A single DL41(DE3)p[KK183] colony was picked into 5ml 2xYT and grown overnight at 37°C.

(2) 700µl of the 5ml 2×YT culture was then added to 300µl sterile 50% glycerol in an Eppendorf tube, and then vortexed.

(3) The mixture was then snap frozen in liquid nitrogen. The resulting glycerol stock was then stored at -70°C.

5.3.4.7 Assessing the Transformation

(1) In order to determine whether the DL41(DE3) cells had been transformed, plasmid DNA was extrated using the QIAGEN miniprep kit (Figure 5.10).

Figure 5.10 Plasmid DNA extraction from the DL41(DE3)p[KK183] expression system.

The extracted plasmid DNA was determined to be about 5Kbp. This corresponds well with the fact that the 4Kbp pKK 183 plasmid has a 1Kbp inserted gene.

(2) The ability of the transformed and untransformed cell strains to grow (overnight at 37°C) on minimal media agar plates containing various supplements was tested (Table 5.20).

Cell Strain	Amino Acid Supplement (60mg/l)	Ampicillin (1μg/ml)	Growth
DL41(DE3)	L-Methionine	×	++++
DL41(DE3)	Seleno-L-methionine	×	++
DL41(DE3)	L-Methionine	✓	-
DL41(DE3)	Seleno-L-methionine	✓	-
DL41(DE3)	×	×	-
DL41(DE3)	×	✓	-
DL41(DE3)p[KK183]	L-Methionine	×	++++
DL41(DE3)p[KK183]	Seleno-L-methionine	×	++
DL41(DE3)p[KK183]	L-Methionine	✓	++++
DL41(DE3)p[KK183]	Seleno-L-methionine	✓	++
DL41(DE3)p[KK183]	×	×	-
DL41(DE3)p[KK183]	×	✓	-

Table 5.20 The supplements required for transformed and untransformed cell growth.

The table shows that both the transformed and untransformed cell strains require methionine or its selenium derivative to grow. This suggests that both strains are methionine auxotrophs. Furthermore, both *E.coli* strains prefer L-methionine over seleno-L-methionine.

The results also show that, unlike the untransformed cell line, DL41(DE3)p[KK183] cells are able to grow on media containing ampicillin.

This proves that the DL41(DE3) strain has indeed been transformed with the pKK 183 plasmid carrying the ampicillin resistance gene.

(3) Single colonies of the various strains were picked into 10ml M9 minimial media containing various supplements. These were then incubated at 37°C and grown for 2 hours. Protein expression in the cells was then induced with 1mM IPTG, and 3 hours later the cells were harvested. D-HicDH activity was then assayed for in the crude cell extracts (Table 5.21).

Cell Strain	Amino Acid Supplement (60mg/l)	Ampicillin (1μg/ml)	O.D.$_{600nm}$	Activity (U)
DL41(DE3)	L-Methionine	×	0.72	0.01
DL41(DE3)	Seleno-L-methionine	×	0.17	0.0
DL41(DE3)p[KK183]	L-Methionine	✓	0.65	46.2
DL41(DE3)p[KK183]	Seleno-L-methionine	✓	0.21	14.3

Table 5.21 D-HicDH activities of the transformed and untransformed cell cells in various growth media.

The results show that there only D-HicDH activity was found in the transformed cells. Moreover, the enzyme activity in the DL41(DE3)p[KK183] cells supplemented with Se-Met was lower than those with the natural methionine substrate. However, it should be noted that the *E.coli* cells growing in seleno-L-methionine also had a lower cell density after 6 hours than the cells grown in methionine.

5.3.4.8 Se-Met D-HicDH Purification Protocol using the DL41(DE3)p[KK183] Methionine Auxotrophic Expression System

After the DL41(DE3)p[KK183] cell growth conditions had been firmly established, a similar purification protocol to the one discussed above (5.3.2.1.6) was followed. This time, however, the 'cocktail' of excess amino acids used to inhibit the methionine biosynthetic pathway was not added. Also, considering that the transformed cells could not synthesise there own methionine, only 0.4% glucose was added to all media (as opposed to 1%). The methionine in the initial starter media would again be more than sufficiently diluted out.

Stage	Volume (ml)	Total protein (mg)	Total activity (U)	Specific activity (U/mg)	Yield (%)	Purification factor -fold
Crude cell extract	34	84	463	5.5	100	1.0
After acid dialysis	43	49	403	8.2	87.0	1.5
After cationic exchange (S - column)	51	34	298	8.8	64.4	1.6
After anionic exchange (Q - column)	73	9.7	214	22.1	46.2	4.0

Table 5.22 Typical D-HicDH selenomethionyl protein purification statistics from a 2 litre methionine auxotrophic DL41(DE3)p[KK183] culture.

Typically, about 6-10mg of pure protein was obtained from a methionine auxotrophic purification preparation. Thus, numerous purification preparations were carried out to obtain sufficient amounts of pure H205Q D-HicDH for crystallisation trials and mass spectrometric analysis.

During the purification procedures their was again still substantial loses in D-HicDH yield when the protein was put down the ionic exchange columns.

The purification statistics above show that the purified D-HicDH had about the same specific activity as that of the pure native enzyme (22.7 U/mg).

This strongly suggests that the incorporation of Se-Met into the protein has very little or no effect on the function of the enzyme.

5.3.4.9 Mass Spectrometric Analysis of Selenomethioyl D-HicDH Protein Produced by using a Methionine Auxotroph

Using ESI-MS, problems were encountered in obtaining a sufficiently concentrated Se-Met protein sample. Therefore, MALDI-MS was employed instead, as this technique does not require a relatively high protein sample concentration. MALDI, however, does need to be calbritrated with a protein of known molecular weight (similar to that of the sample to be measured). Again, problems were encountered in trying to get a decent spectrum from the known native H205Q D-HicDH protein sample. Numerous attempts to desalt and concentrate the native sample failed. Thus, the commercially available aldolase enzyme from *E.coli*, which has a mass of 37979.3Da was used to calibrate the MALDI machine. After further desalting with a C18 chromatographic media loaded pippette tip (Waters), the mass of theSe-Met H205Q D-HicDH protein sample was measured (Figure 5.11).

The spectrum suggests that aldolase was not sufficiently similar in mass to the D-HicDH protein to produce an accurate, clear-cut, single mass peak. However, the spectrum does suggest that there is probably only one protein ion species in solution with a mass of about 37,123 Daltons; this peak mass corresponds to the H205Q D-HicDH ion having its N-terminal methionine removed with the remaining 9 methionine residues all being substituted for Se-Met residues (*cf.* Figures 5.8 and 5.11). This is supported by the fact that the overwhelming concentration of substrate in the growth media was in the form of Se-Met (about 99.9999%), and coupled with the fact that the host

strain was unable to synthesise methionine, this strongly suggests that Se-Met was fully incorporated into H205Q D-HicDH.

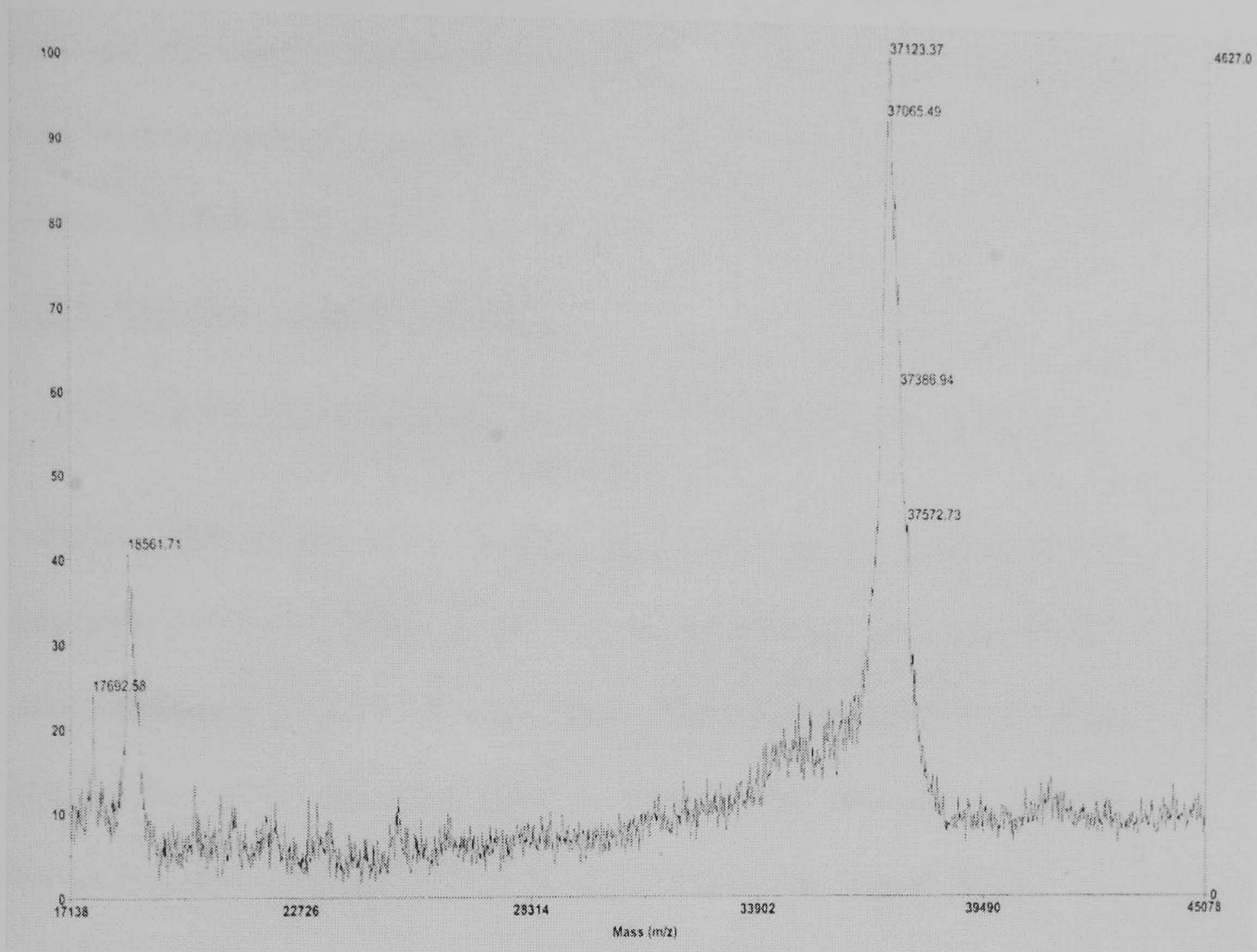


Figure 5.11 MALDI mass spectrum of a selenomethionyl H205Q D-HicDH protein sample produced by a methionine auxotroph.

5.3.4.10 Crystallising the Selenomethionyl H205Q D-HicDH Protein Produced using a Methionine Auxotroph

The crystallisation conditions for selenomethionyl proteins are very similar to those for the natural proteins (Drenth, 1995). Se-Met protein crystals are usually grown in the presence of DTT and EDTA to avoid oxidation.

The crystals that form tend to be smaller and usually diffract to a lower resolution compared with their native counterparts (Doublié, 1997).

Many binary Se-Met H205Q crystals of type (3) (*i.e.* 35-40mg/ml protein, 1mM NADH; 20-25% PEG MME 5000, 0.2M ammonium sulphate, 0.1M

MES pH 6.5) were grown in the presence of 0.2mM EDTA and 5mM DTT, as these reproducible crystals gave the best diffraction.

Although they shared the same morphology, the Se-Met crystals that grew were slightly smaller than the corresponding native crystals. These crystals were then cryo-cooled and stored in liquid nitrogen as soon as possible again to avoid oxidation (Hendrickson *et al.*,1990)

5.3.5 Attempts at MAD phasing

5.3.5.1 X-ray Fluorescence: Choosing Wavelengths for MAD Data Collection

Before starting the MAD phasing experiment at the Daresbury SRS laboratory (PX9.5), a X-ray fluoresence experiment was performed using a selenomethionyl D-HicDH protein crystal. The program *CHOOCH* (Evans and Pettifer, 2001) was then used to calculate the anomalous scattering factors from the subsequent X-ray fluoresence spectrum (Figure 5.12).

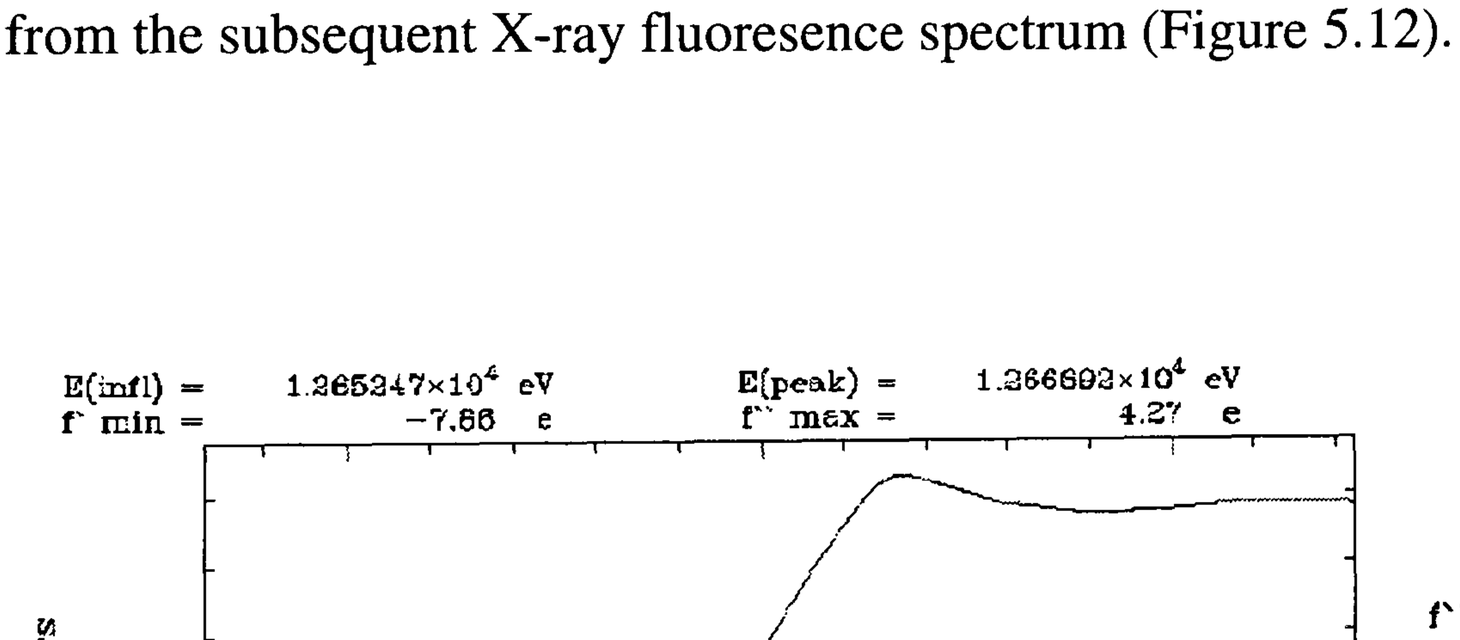


Figure 5.12 Plot showing experimental values for f'' and f' as a function of X-ray energy. (NB. 1KeV = 12.398 / λ in Å).

The plot above clearly indicates that the Se-Met D-HicDH crystals should be suitable for MAD phasing experiments.

5.3.5.2 Attempts at Data Collection at the Absorption Edge

In all 17 binary Se-Met H205Q crystals of type (3) ($P2_1$ and C2) were mounted. Each crystal diffracted very poorly (5-6Å) and had a large mosaic spread (~1.5°). These crystals were not adequate for MAD as this technique requires crystals to diffract to at least 3.5Å and have a small R_{merge} (to determine the background noise in measuring Bijvot differences) in order for useful phase information to be extracted.

5.3.6 Conclusions

Other crystal forms of the Se-Met protein were grown such as the wild type D-HicDH apoenzyme type (1) crystals and the H205Q mutant binary enzyme type (2) crystals (see Chapter 4).

When taken up to Daresbury (PX 9.5). The alternative crystal complexes and forms were also found to diffract poorly and thus could not be used in MAD phasing.

It seems as though the poor diffraction of the original native crystals was made even worse by the incorporation of selenomethionine into the protein.

Chapter 6

Subunit Interface Mutants of L-Lactate Dehydrogenase from the Malarial Parasite *Plasmodium Falciparum*

6.1 Malaria

6.1.1 Introduction

Malaria is an infectious disease which is estimated to kill between 1.5-2.7 million people annually (WHO, 1996), the majority of these being children under the age of five (Knell, 1991). Another 300 to 500 million people have the disease.

A third of the world's population, ~1.78 billion, live in zones where they are at risk of catching malaria. Countries in tropical Africa account for more than 90 percent of the cases and more than 6 percent occur in India, Brazil, Sri Lanka, Afghanistan, Vietnam and Colombia (WHO, 1987).

6.1.2 The Causes of Malaria

Malaria is transmitted by the parasitic protozoa of the genus *Plasmodium*.

The female mosquitoes of the genus *Anopheles* are the sole hosts of human malarial parasites (Knell, 1991). Whilst feeding on blood, the mosquitoes transmit the *Plasmodia* to humans through saliva. On entering a human, the parasite invades a liver cell where it rapidly multiplies. Eventually, the liver cell ruptures and releases the parasites into the bloodstream where they infect red blood cells. Most parasites again reproduce, killing the red blood cells and then invading more blood cells. Other parasites, while in the red blood cells, develop into male and female form. When these cells are ingested by a feeding mosquito, the cells burst, freeing the sexual forms of the parasite.

The two forms merge creating an oocyst; after maturing in the stomach wall of the mosquito, the oocyst ruptures to release thousands of parasites. These then migrate to the mosquito's salivary glands. When the insect feeds on another human, saliva containing the parasite is injected into the blood and therefore the infective cycle can start again.

There are four species of *Plasmodia* that infect humans: *Plasmodium vivax*, *Plasmodium ovale*, *Plasmodium malariae* and *Plasmodium falciparum*.

The first three may cause severe illness but are rarely fatal. However, the fourth type of human parasite, *Plasmodium falciparum*, is much more serious and often leads to coma and death within a few days.

6.1.3 Combating Malaria

The life cycles of the *Plasmodium* parasite and humans are linked by the *Anopheles* vector. Initial attempts at eradicating malaria aimed to break this link, thus causing *Plasmodium* to perish.

In 1942, the insecticide dichlorodiphenyl-trichloroethane (DDT) was discovered. DDT controlled the population of mosquitoes and thus reduced the outbreak of the malarial disease. However, despite its initial success, DDT failed to eradicate the disease. This failure was, in part, due to vector and parasite drug resistance (WHO, 1987).

Recent malarial drug design strategies have focused on attacking the parasite; The main drugs in general use during the last century can be divided into two categories.

The first class of anti-malarial drugs are the antifolates. These interrupt the synthesis of DNA in the parasite and are represented by the dihydrofolate reductase (DHFR) inhibitors, such as pyrimethamine and proguinal, and the sulfa drugs.

Unfortunately, antifolate resistance in both *P.falciparum* and *P.vivax* developed rapidly following the introduction of pyrimethamine in many parts of the tropics (Knell, 1991).

The mode(s) of action of the second class of anti-malarials is not fully understood. It is believed that the compounds form complexes with haematin, inducing aggregrate formation in the parasitic cell (Carroll and Behrens, 1998). This class of drugs includes the compounds quinine, chloroquine and mefloquine (Figure 6.1).

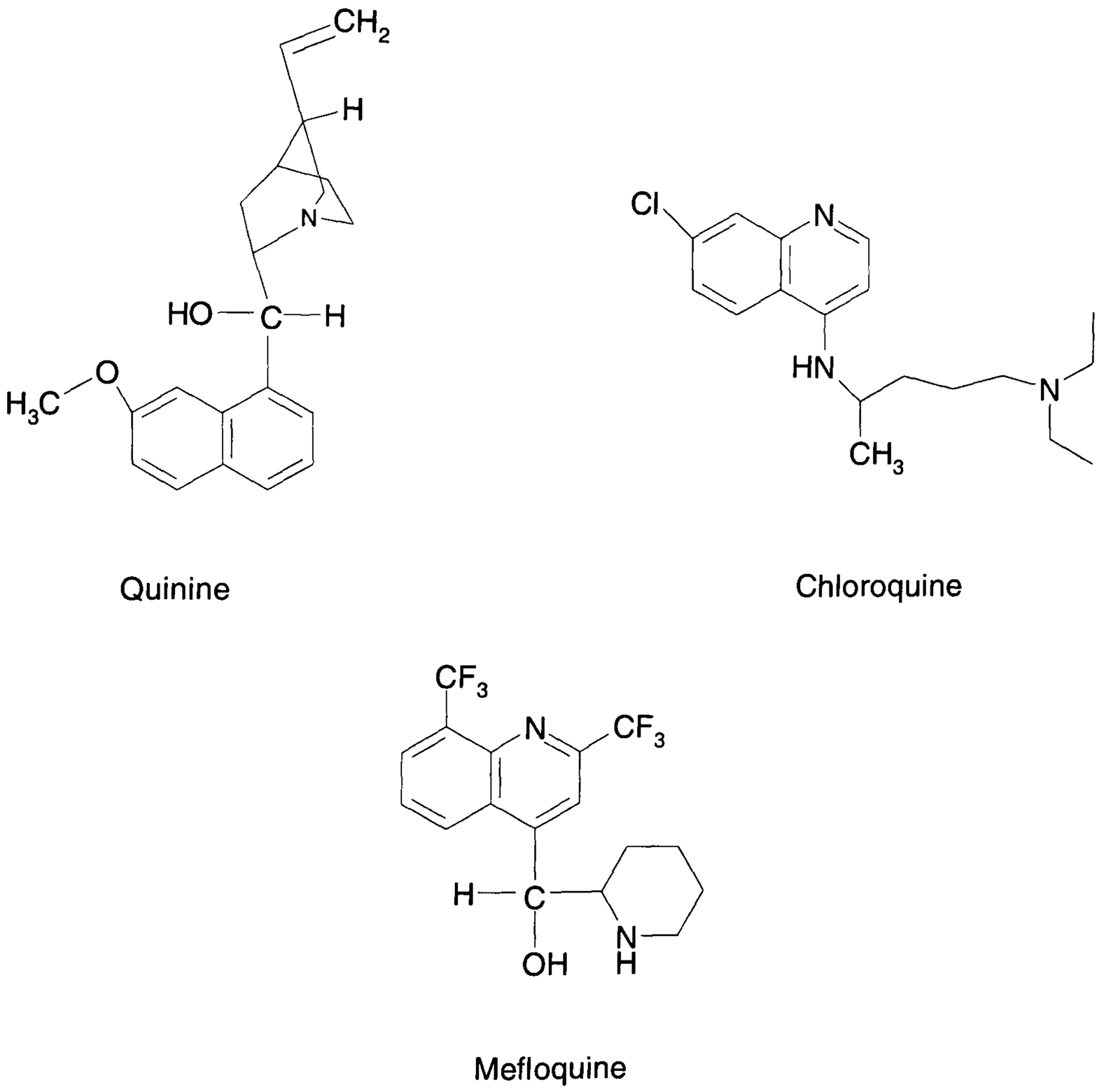


Figure 6.1 The chemical structures of quinine, chloroquine and mefloquine.

Chloroquine is the most widely used anti-malarial drug. Aside from adverse side effects related to the use of chloroquine, resistance to the compound (and other quinoline containing drugs) is also increasing at an alarming rate in all regions where malaria is endemic (Knell, 1991).

The resistance of the malarial parasite, especially *P.falciparum*, to existing anti-malarial drugs is a serious and increasing problem in many parts of the world and the need for novel antimalarial drugs is great. The majority of drugs in use today act by inhibiting parasitic metabolic processes which are different from those of the host. Therefore in the search for new drugs, metabolic pathways or processes which are specific or sufficiently different from that of the host, are often the targets for structure-based drug design.

6.2 *Plasmodium falciparum* L-Lactate Dehydrogenase

Design of an effective antimalarial requires specific targeting of the metabolic pathways vital to the parasites' survival. The enzymes of the glycolytic pathway have been observed in *Plasmodium*, and identified both as potential antimalarial drug targets and indicators of blood parasitaemia levels in diagnosis (Makler and Hinrichs, 1993; Gomez *et al.*,1997; Olliaro and Yuthavong, 1999).

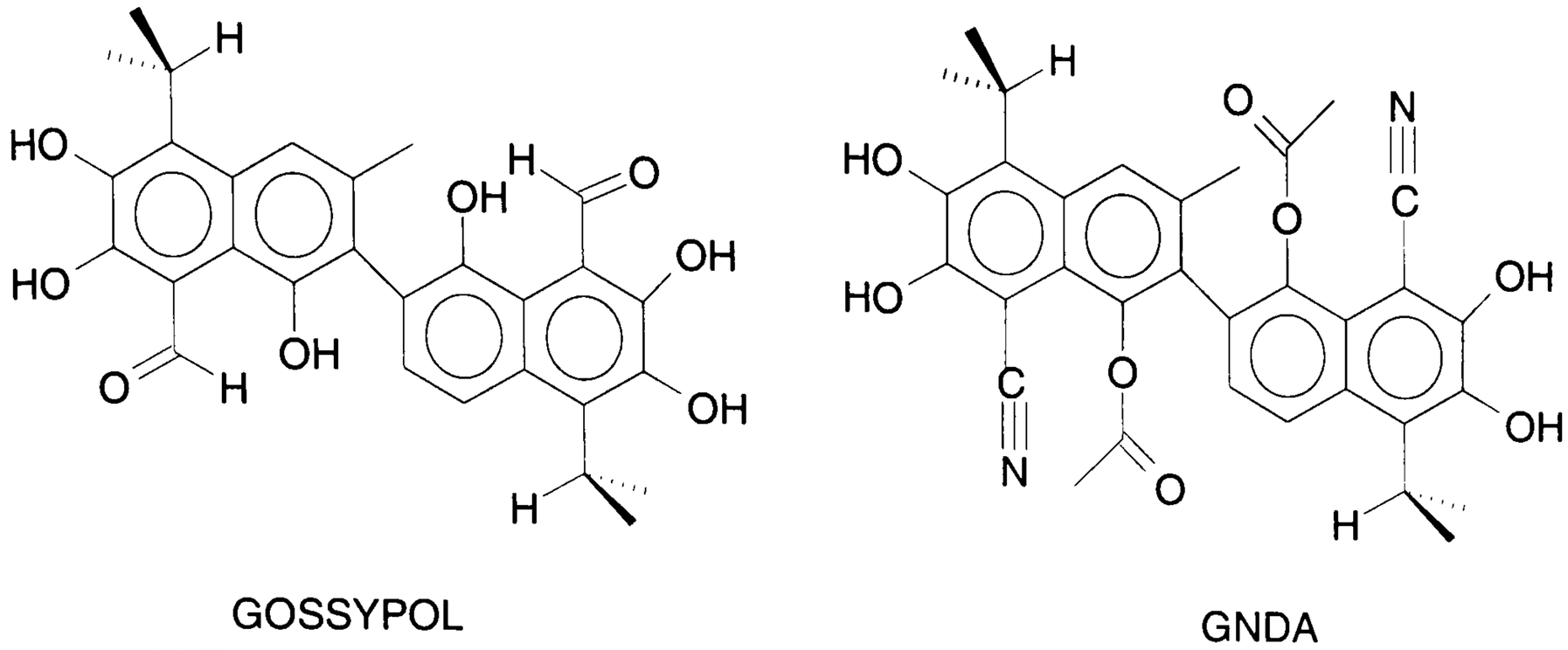
The L-LDH enzyme from *Plasmodium falciparum* (*Pf*LDH) has been shown to be electrophoretically, kinetically and structurally distinct from mammalian forms of the enzyme, including human (Vander Jagt *et al.*, 1981; Dunn *et al.*, 1996).

Malarial parasites depend on anaerobic glycolysis to satisfy much of their ATP demands (Vander Jagt *et al.*, 1981). They require a high level of energy production to maintain rapid multiplication rates during the asexual, intraerythrocytic cycle in the human host. Royer *et al.* (1986) demonstrated that molecules which inhibit *Pf*LDH are lethal to the parasite. Therefore, this enzyme is considered suitable as a potential antimalarial drug target (Sessions *et al.*, 1997).

The crystal structure of the ternary complex of the *Pf*LDH (enzyme + NADH + oxamate) revealed a number of differences compared with other known L-LDH structures (Dunn *et al.*, 1996).

Firstly, the cofactor is significantly displaced in *Pf*LDH: The shift is more prominent at the nicotinamide end of NADH (closest to the active site), with the equivalent positions of the atoms in the nicotinamide ring differing by an average of 1.2Å between the mammalian and malarial structures. This displacement of the NADH may explain why *Pf*LDH, unlike other known L-LDHs, is not inhibited by high levels of substrate (presumably as an adaption to the pyruvate-rich conditions of the red blood cell): The altered position of the coenzyme may cause the NAD^+ off-rate to increase, thus reducing the formation of the NAD^+-pyruvate inhibitory adduct.

Secondly, *Pf*LDH differs from other known L-LDHs in that it contains a 5 amino acid insert in the substrate specificity loop, linking α-D/E to β-E, (Figure 1.4; *cf. Pf*LDH and *L.confusus* L-HicDH). A distinctive cleft is formed alongside this extended loop, this contrasts sharply with the shallow groove found in the equilvalent postion in mammalian L-LDHs. It has been known for sometime that L-LDHs are inhibited by derivatives of gossypol (Royer *et al.*, 1986), and that these compounds bind close to the pyruvate/nicotinamide-ribose site (Sessions *et al.*, 1997). However, the gossypol derivative gossylic nitrile diacetate (GNDA) selectively inhibits *Pf*LDH compared to the human enzyme (Sessions *et al.*, 1997) (Figure 6.2).



GOSSYPOL GNDA

Figure 6.2 The chemical structures of gossypol and GNDA.

This difference in binding may be attributed to the ability of the malarial enzyme to accomodate the acetyl groups of GNDA in its deep cleft. Thus, there is potential to use this cleft to provide selectivity in the rational structure-based drug design of antimalarial.

Also, recently a crystal structure complex of *Pf*LDH and chloroquine has been reported (Read *et al.*, 1999). Chloroquine binds in a similar postion to that of the adenyl ring of the cofactor and acts a modest competitive inhibitor. Even though the mechanism by which chloroquine exerts its effects on *P.falciparum* remains unclear, the ligand can be used to act as a template for new chemical compounds that specifically and efficiently inhibit *Pf*LDH.

6.3 Mutations at the Subunit Interface of *Pf*LDH

A substational amount of the *Pf*LDH protein is required for malarial drug design / selectivity studies. So, in order to obtain a reproducible purification protocol for the enzyme a non-cleavable 6 Histidine tag had been attached at the C-terminus of the protein; furthermore it was shown that the 'His-tag' has very little effect on the kinetic behaviour of *Pf*LDH (Shoemark, 2000). However, it was also shown that, unless a high percentage of β-mercaptoetanol was added, the His-tagged enzyme would appear to run as a dimer at high concentrations on denaturating SDS PAGE. Furthermore,

kinetic experiments have also demonstrated that there is a loss in activity if the His-tagged enzyme was oxidised; and this activity could be recovered by the addition of dithiothreitol.

It was thought that these unusual effects may have been the result of the formation of inappropriate disulphide bonds between two cysteine residues at the subunit interface. Indeed, the crystal structure revealed that residues C183 and C76 came into apposition at the dimer subunit interface of the wild type *Pf*LDH (Figure 7.17).

Of the two residues, C183 appeared to have the fewer space constraints exerted by the surrounding polypeptide. Therefore the cysteine 183 residue was substitued by valine, alanine and serine, and the atmospheric stability of the resulting mutants evaluated by Shoemark (2000):-

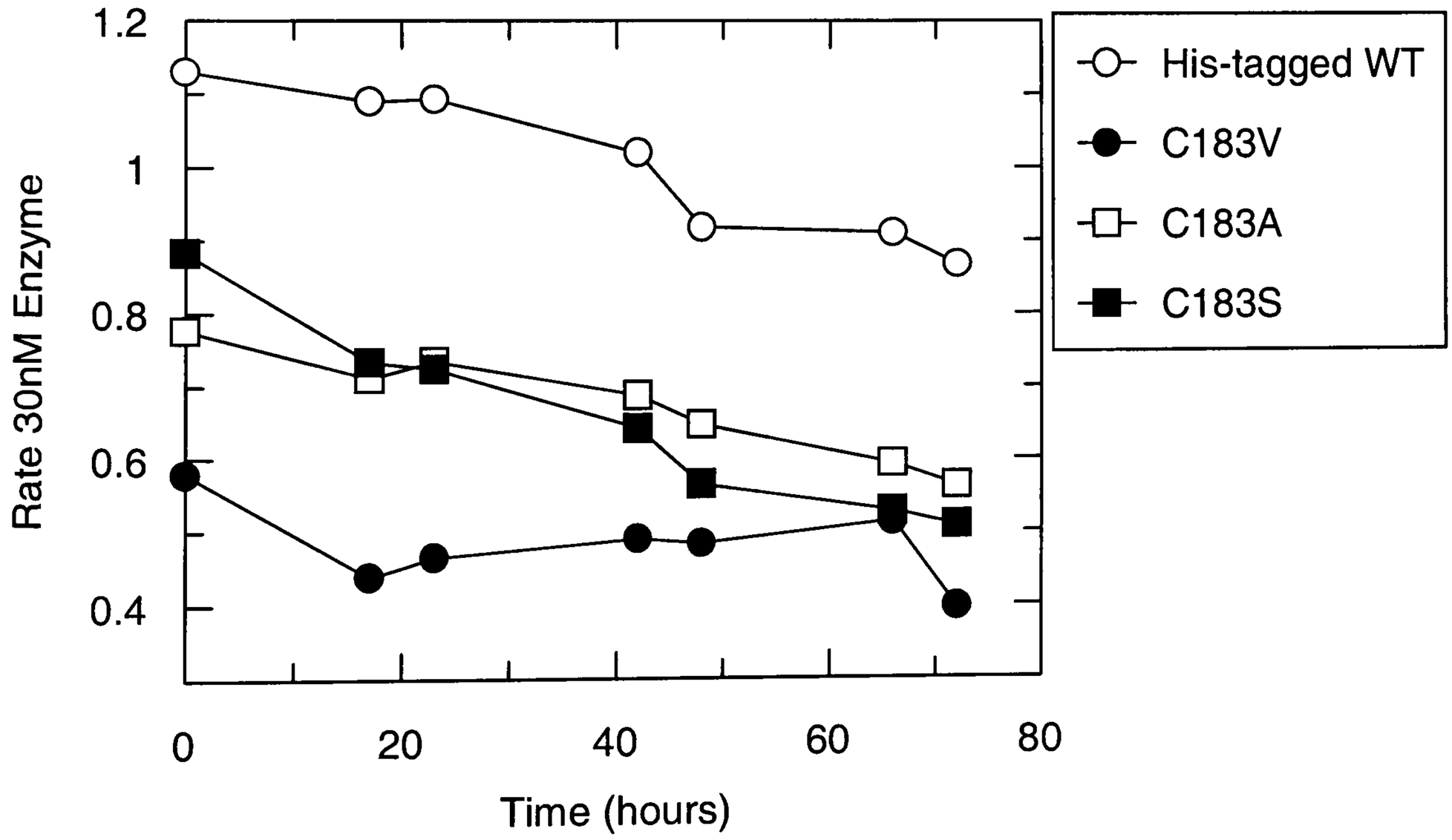


Figure 6.3 Atmospheric Stability of the C183 mutants compared with the wild type His-tagged *Pf*LDH enzyme. The samples were incubated at 25°C in a buffer containing 50mM Tris and 50mM KCl at pH 7.5. A loosely applied lid was screwed onto the sample tubes.

The results show that all the C183 mutants had enzymatic rates slightly lower than the wild type enzyme.

Surprisingly however, the C183A and C183S mutants were not more stable against oxidation than the wild type *Pf*LDH; in fact the rate of loss of activity for the two mutants was very similar to that for the wild type enzyme.

Only the C183V mutant appeared to be slightly more stable against oxidation when compared with the wild type enzyme. However, this was at the cost of a reduction in k_{cat} of about 40%.

Chapter 7

Determination and Analysis of the Molecular Structures of the C183 *Pf*LDH Subunit Interface Mutants

7.1 Aims

Frozen purified protein samples (stored in 10% glycerol at -70°C) were obtained from Deborah Shoemark in order to determine if there were any structural pertubations at the subunit interfaces of the mutant enzymes, and whether they were the cause of the modified atmospheric stabilites in the C183 *Pf*LDH mutants.

7.2 Crystallisation of the Subunit Interface Mutant *Pf*LDH Ternary Complexes

Any bound cofactor was removed from the protein by dialysing each of the thawed samples against 0.5g/l activated charcoal until the A_{280nm}/A_{260nm} measured above 1.9. The samples were then concentrated and buffer exchanged against milli Q water using 30kD MWCO Centricon filters. The extinction coefficient at 280nm, for a 1% solution of *Pf*LDH was taken to be 0.47 (Dunn *et al.*, 1996).

Crystals of all 3 *Pf*LDH mutants of the ternary complex form were obtained using a method similar to one previously devised by Jonathon Read (unpublished results) for the wild type His-tagged enzyme.

Crystals grew at 17°C using the hanging drop vapour diffusion method against a well containing 24-30% PEG 1,500. The mother liqour contained 24-32 mg/ml protein, 200mM imidazole, 50mM Hepes pH 7.5, 1mM NADH and 1mM sodium oxamate.

Oxamate is an iso-steric and iso-electronic inactive substrate analogue of pyruvate which forms a stable abortive ternary complex with NADH in lactate dehydrogenases (Wigley *et al.*, 1992; McPherson, 1973).

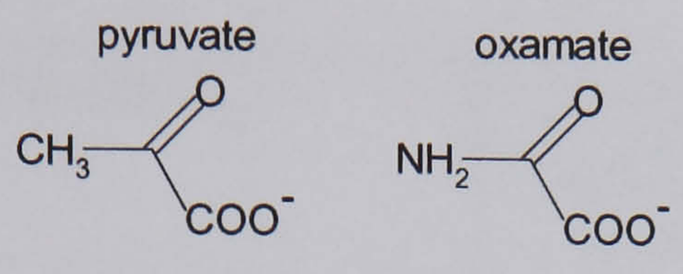


Figure 7.1 The chemical structures of pyruvate and oxamate.

The ternary complex of the C183 mutants was found to crystallise as clusters of long rectangular rods. For data collection purposes, the clusters were separated into single crystals by micro-manipulation (Figure 7.2).

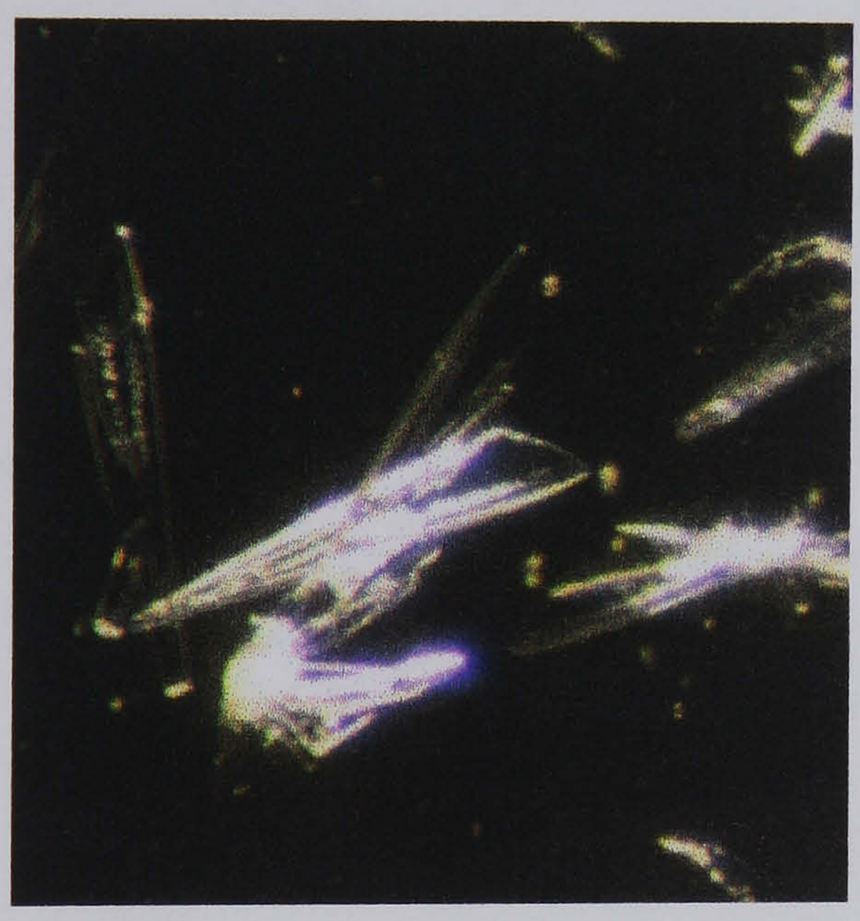


Figure 7.2 Typical ternary complex crystals of the *Pf*LDH C183 mutants.

7.3 Data Collection

Diffraction data for the C183V mutant *Pf*LDH were collected from a single crystal at the Daresbury-SRS laboratory (PX 7.2, λ=1.488Å) at a temperature of 100K. The data statistics are summarised in Table 7.1.

The in-house rotating anode X-ray source (λ=1.5418Å) was used to collect diffraction data, at 100 K, for both the C183A and C183S mutants from single crystals. The diffraction data statistics for the C183A and C183S proteins are summarised in Tables 7.2 and 7.3, respectively.

Data collected using the Daresbury SRS laboratory (PX7.2, λ = 1.488Å).

Cryoprotectant solution	32% PEG 1,500, 200mM Imidazole, 50mM Hepes pH 7.5, 1mM NADH, 1mM oxamate, 10% glycerol.
Unit cell	a=80.38, b=86.16, c=90.19, $\alpha=\beta=\gamma=90^{\circ}$
Space group	I222
Total number of unique reflections	18,239
Mean Redundancy	4.71
Resolution	2.1Å
Mosaicity	1.36°
Completeness - Overall (30.00-2.1Å) Outer shell (2.07-2.1Å)	97.4% 96.3%
I / σ (I) - Overall Outer shell	13.14 4.00
R_{merge} - Overall Outer shell	9.4% 26.0%

Table 7.1 Diffraction data collection statistics of the ternary C183V *Pf*LDH mutant crystal.

$V_M = 2.3$ Å^3/Da $\Rightarrow$ 1 molecule/a.s.u.

Data collected using the in-house rotating-anode X-ray source (λ = 1.5418Å).

Cryoprotectant solution	30% PEG 1,500, 200mM Imidazole, 50mM Hepes pH 7.5, 1mM NADH, 1mM oxamate, 10% glycerol.
Unit cell	a=80.42, b=86.76, c=90.62, $\alpha=\beta=\gamma=90°$
Space group	I222
Total number of unique reflections	35,072
Mean Redundancy	3.84
Resolution	1.7Å
Mosaicity	0.45°
Completeness - Overall (22.00-1.7Å) Outer shell (1.76-1.7Å)	99.7% 99.9%
I / σ (I) - Overall Outer shell	10.58 3.19
R_{merge} - Overall Outer shell	10.8% 34.8%

Table 7.2 Diffraction data collection statistics of the ternary C183A *Pf*LDH mutant crystal.

$V_M = 2.3$ Å^3/Da $\Rightarrow$ 1 molecule/a.s.u.

Data collected using the in-house rotating-anode X-ray source (λ = 1.5418Å).

Cryoprotectant solution	30% PEG 1,500, 200mM Imidazole, 50mM Hepes pH 7.5, 1mM NADH, 1mM oxamate, 10% glycerol.
Unit cell	a=80.49, b=86.79, c=90.80, $\alpha=\beta=\gamma=90°$
Space group	I222
Total number of unique reflections	25,159
Mean Redundancy	4.30
Resolution	1.9Å
Mosaicity	0.76°
Completeness - Overall (22.00-1.9Å) Outer shell (1.97-1.9Å)	98.4% 94.5%
I / σ (I) - Overall Outer shell	8.22 3.11
R_{merge} - Overall Outer shell	18.4% 52.8%

Table 7.3 Diffraction data collection statistics of the ternary C183S *Pf*LDH mutant crystal.

(The R_{merge} is particularly high in the outer resolution shell).

$V_M = 2.3$ Å^3/Da $\Rightarrow$ 1 molecule/a.s.u.

Analytical ultra-centrifugation had previously been used to show that *Pf*LDH was in a tetrameric state in solution (Shoemark, 2000). And it was found that all three *Pf*LDH mutants crystallised in space group I222, with one monomer in the asymmetric unit. Therefore the three molecular two-fold axis, P, Q and R (described by Rossmann *et al.*, 1971) must be coincident with the crystallographic axes.

7.4 Solving the C183 Mutant *Pf*LDH Structures

7.4.1 Introduction

After each of the data sets had been processed and reduced, the structures were solved using molecular replacement. A monomer of the original, 'non-His-tagged', wild type *Pf*LDH:NADH:oxamate ternary complex was used as a search model in all three cases (Dunn *et al.*, 1996, structure at 1.74Å resolution, PDB I.D.: 1LDG).

The PDB header information, solvent molecules, and the NADH and oxamate ligands were removed from the model's coordinates file before running *AMoRe*.

In each case, the top 20 rotation solutions were put into the translation function. Subsequently, the clear translation peak underwent rigid body refinement in the 'fiting' function in order to improve the solution.

7.4.2 Solving the C183V Mutant

Radius of integration: 25Å. Cell model: 125Å × 125Å × 125Å.

Resolution cutoff: 10.0-4.0Å.

Solution Number	**Rotation of Eulerian angles (°)**			**Correlation coefficient**
	α	β	γ	(%)
Solution 1	98.73	56.71	230.73	26.7
Solution 2	2.24	88.57	206.31	10.2
Solution 3	110.74	85.11	192.67	10.1
Solution 4	177.81	90.00	28.65	10.0
Solution 5	172.84	59.30	126.13	10.0
Solution 6	94.00	29.56	223.50	9.9
Solution 7	66.72	39.98	28.01	9.8
Solution 8	44.61	30.31	59.95	9.7
Solution 9	171.71	72.37	325.70	9.6
Solution 10	162.22	75.53	319.00	9.6

Table 7.4 The rotation solutions from *AMoRe* for the ternary C183V *Pf*LDH mutant complex using the monomer of the wild type *Pf*LDH as the search model.

Solution Number	**Rotation of Eulerian angles (°)**			**Translation (fractions of the unit cell)**			**Correlation coefficient**	**R-factor**
	α	β	γ	**x**	**y**	**z**	(%)	(%)
Solution 1	98.73	56.71	230.73	0.3210	0.3105	0.0989	70.7	33.2
Solution 2	2.24	88.57	206.31	0.0486	0.4577	0.4178	13.5	54.2
Solution 3	110.74	85.11	192.67	0.1988	0.4385	0.0546	14.4	53.5
Solution 4	177.81	90.00	28.65	0.2469	0.3867	0.3312	14.9	53.4
Solution 5	172.84	59.30	126.13	0.0199	0.2468	0.2811	15.1	53.6
Solution 6	94.00	29.56	223.50	0.0069	0.0103	0.4256	15.8	53.8
Solution 7	66.72	39.98	28.01	0.3812	0.1618	0.2071	12.9	54.2
Solution 8	44.61	30.31	59.95	0.1183	0.0511	0.0808	13.9	53.2
Solution 9	171.71	72.37	325.70	0.2665	0.3098	0.0213	12.7	54.5
Solution 10	162.22	75.53	319.00	0.2859	0.0731	0.1220	13.8	53.9

Table 7.5 The translation solutions from *AMoRe* for the ternary C183V *Pf*LDH mutant complex using the monomer of the wild type *Pf*LDH as the search model.

Solution Number	**Rotation of Eulerian angles (°)**			**Translation (fractions of the unit cell)**			**Correlation coefficient**	**R-factor**
	α	β	γ	**x**	**y**	**z**	(%)	(%)
Solution 1	99.93	56.78	230.43	0.3219	0.3105	0.0986	76.3	29.1

Table 7.6 The final solution after 'fiting' from *AMoRe* for the ternary C183V *Pf*LDH mutant complex using the monomer of the wild type *Pf*LDH as the search model.

7.4.3 Solving the C183A Mutant

Radius of integration: 25Å. Cell model: 125Å × 125Å × 125Å.

Resolution cutoff: 10.0-4.0Å.

Solution Number	**Rotation of Eulerian angles (°)**			**Correlation coefficient (%)**
	α	β	γ	
Solution 1	99.28	56.69	230.53	27.6
Solution 2	107.26	14.68	20.18	10.5
Solution 3	47.54	80.19	291.50	10.3
Solution 4	2.75	87.06	205.50	9.9
Solution 5	171.14	59.16	126.50	9.8
Solution 6	162.27	60.77	239.00	9.2
Solution 7	51.50	27.07	136.00	9.2
Solution 8	20.60	62.55	186.97	9.1
Solution 9	3.52	57.93	111.33	9.0
Solution 10	36.46	58.92	229.70	8.9

Table 7.7 The rotation solutions from *AMoRe* for the ternary C183A *Pf*LDH mutant complex using the monomer of the wild type *Pf*LDH as the search model.

Solution Number	**Rotation of Eulerian angles (°)**			**Translation (fractions of the unit cell)**			**Correlation coefficient**	**R-factor**
	α	β	γ	**x**	**y**	**z**	(%)	(%)
Solution 1	99.28	56.69	230.53	0.3210	0.3108	0.0981	72.8	31.2
Solution 2	107.26	14.68	20.18	0.3650	0.1208	0.3437	14.5	52.6
Solution 3	47.54	80.19	291.50	0.4276	0.4261	0.3466	12.2	53.0
Solution 4	2.75	87.06	205.50	0.4273	0.4313	0.1965	14.2	52.0
Solution 5	171.14	59.16	126.50	0.0393	0.0444	0.4659	15.2	52.8
Solution 6	162.27	60.77	239.00	0.0079	0.4593	0.4086	13.1	52.9
Solution 7	51.50	27.07	136.00	0.2835	0.0906	0.4495	13.1	52.9
Solution 8	20.60	62.55	186.97	0.3934	0.0948	0.4583	13.0	52.1
Solution 9	3.52	57.93	111.33	0.0007	0.2939	0.4616	13.1	52.8
Solution 10	36.46	58.92	229.70	0.0403	0.2608	0.1371	12.7	52.2

Table 7.8 The translation solutions from *AMoRe* for the ternary C183A *Pf*LDH mutant complex using the monomer of the wild type *Pf*LDH as the search model.

Solution Number	**Rotation of Eulerian angles (°)**			**Translation (fractions of the unit cell)**			**Correlation coefficient**	**R-factor**
	α	β	γ	**x**	**y**	**z**	(%)	(%)
Solution 1	99.89	56.63	230.25	0.3221	0.3108	0.0983	77.9	27.5

Table 7.9 The final solution after 'fiting' from *AMoRe* for the ternary C183A *Pf*LDH mutant complex using the monomer of the wild type *Pf*LDH as the search model.

7.4.4 Solving the C183S Mutant

Radius of integration: 25Å. Cell model: 125Å × 125Å × 125Å.

Resolution cutoff: 10.0-4.0Å.

Solution Number	**Rotation of Eulerian angles (°)**			**Correlation coefficient (%)**
	α	β	γ	
Solution 1	99.05	56.71	230.64	29.4
Solution 2	2.64	87.42	206.16	11.3
Solution 3	107.17	14.70	20.18	11.3
Solution 4	177.85	90.00	29.00	10.7
Solution 5	172.22	58.95	127.09	9.6
Solution 6	4.42	57.72	110.71	9.2
Solution 7	14.15	82.45	88.21	9.1
Solution 8	48.4	79.72	291.41	9.0
Solution 9	94.63	29.98	224.00	8.8
Solution 10	4.55	38.78	327.44	8.8

Table 7.10 The rotation solutions from *AMoRe* for the ternary C183S *Pf*LDH mutant complex using the monomer of the wild type *Pf*LDH as the search model.

Solution Number	**Rotation of Eulerian angles (°)**			**Translation (fractions of the unit cell)**			**Correlation coefficient**	**R-factor**
	α	β	γ	**x**	**y**	**z**	(%)	(%)
Solution 1	99.05	56.71	230.64	0.3210	0.3110	0.0980	74.2	30.9
Solution 2	2.64	87.42	206.16	0.2690	0.4951	0.0549	14.1	53.3
Solution 3	107.17	14.70	20.18	0.3651	0.1204	0.3429	14.3	53.7
Solution 4	177.85	90.00	29.00	0.2404	0.3880	0.3375	14.6	52.8
Solution 5	172.22	58.95	127.09	0.1879	0.4208	0.2021	15.0	52.8
Solution 6	4.42	57.72	110.71	0.2355	0.1041	0.0354	12.7	53.8
Solution 7	14.15	82.45	88.21	0.3181	0.1954	0.4639	14.3	53.5
Solution 8	48.4	79.72	291.41	0.4159	0.4618	0.0088	11.5	54.2
Solution 9	94.63	29.98	224.00	0.4371	0.3792	0.2472	15.2	52.6
Solution 10	4.55	38.78	327.44	0.2174	0.1690	0.1991	12.8	53.5

Table 7.11 The translation solutions from *AMoRe* for the ternary C183S *Pf*LDH mutant complex using the monomer of the wild type *Pf*LDH as the search model.

Solution Number	**Rotation of Eulerian angles (°)**			**Translation (fractions of the unit cell)**			**Correlation coefficient**	**R-factor**
	α	β	γ	**x**	**y**	**z**	(%)	(%)
Solution 1	99.79	56.72	230.33	0.3221	0.3109	0.0983	78.8	27.5

Table 7.12 The final solution after 'fiting' from *AMoRe* for the ternary C183S *Pf*LDH mutant complex using the monomer of the wild type *Pf*LDH as the search model.

The final solutions for all three mutants have very high correlation coefficients and low R-factor values.

For each C183 *Pf*LDH mutant, the coordinates of the wild-type PfLDH search model were rotated and translated to the position in the unit cell determined by molecular replacement using the CCP4 *lsqkab* program.

7.5 Crystallographic Refinement and Model Building of the C183 Mutant *Pf*LDH Structures

Prior to refinement, an R_{free} test set was generated for each of the three mutants using the CCP4 program *unique*. The data sets contained 5% of the unique reflections (greater than 500 reflections), as required for statistical validity of the figure (Brünger, 1992b).

All three structures were refined using similar methods (using all the available data).

The structures were all initally subjected to twenty cycles of refinement using the *Refmac* CCP4 program (Murshudov *et al.*, 1999), using the maximum-likelihood option with overall B-factor refinement.

After converting the mtz reflection files into the appropriate cns format using *mtz2various* (CCP4), refinement of the structures was then continued using the CNS suite of programs (Crystallography and NMR System) (Brünger *et al.*, 1998).

After initially resetting each B-factor to 15.0Å^2, a round of simulated annealing (starting temperature 2500K) was run. In CNS the simulated annealing refinement task includes conjugate-gradient energy minimisation of the coordinates both before and after the simulated annealing; 200 cycles of minimisation were run in both instances. This was followed by individual B-factor refinement.

A second round of simulated annealing (at a starting temperature of 2500K) and conjugate-gradient minimisation (200 cycles each) were then run. This was followed by searching for solvent sites in $|F_{obs}|-|F_{calc}|$ difference maps at positive sites with a sigma level cutoff greater than 3.0, with temperature factors less than 50Å^2 and within hydrogen bonding distance (3.5Å) (using the *water_pick* program in CNS). These sites were modelled as water molecules. The *water_pick* task includes 20 cycles of minimisation of the coordinates followed by 20 cycles of minimisation of the B-factors.

$|F_{obs}|-|F_{calc}|$ and $2|F_{obs}|-|F_{calc}|$ difference maps were then generated and converted to the omap format (using *cns2omap*) and the subsequent electron density maps, along with the coordinate files, viewed using the graphics program *O* (Jones *et al.*, 1991). Any water molecules that appeared positionally unstable (*i.e.* in poor electron density, near hydrophobic side chains, *etc*) were removed.

All three C183 *Pf*LDH mutant maps had positive densities indicating the presence of a NADH molecule and an oxamate molecule, these were then subsequently built into the maps (Figure 7.3). Moreover, positive density corresponding to a glycerol molecule was found in exactly the same position in all three mutant maps, this molecule was also built in to the structures (Figure 7.19).

100 cycles of energy minimisation were then run, followed by individual B-factor refinement. At this stage further water molecule sites were identified in $|F_{obs}|-|F_{calc}|$ and difference maps at positive sites > 3σ and a B-factor less than 50Å^2.

$|F_{obs}|-|F_{calc}|$, $2|F_{obs}|-|F_{calc}|$ and $3|F_{obs}|-2|F_{calc}|$ electron density maps were then generated and the structures rebuilt (and ligands checked) with *O*. The *rsr*-

rotamer and *lego-side-chain* commands in *O* were used to identify the amino acid sidechain conformations that fitted the electron density the best. Also, a macro was read into *O* that enabled the water molecules in the structures to be located and thus checked to see if they were in density. The occupancy of disordered parts of sidechains was set to zero.

In all three mutant structures it was clearly evident that cysteine 183 (a169 in the PDB files) had been mutated. Furthermore it was obvious that the positive and negative regions of electron density around this amino acid corresponded precisely with the proposed mutations, the appropriate amino acids were then substituted into the structure (Figures 7.4-7.6). The mutagenesis was later confirmed by sequencing the DNA of the C183V, C183A and C183S *Pf*LDHs (Shoemark, 2000).

100 cycles of conjugate-gradient minimisation were then followed by individual B-factor refinement, which in turn was followed by a search for more water molecules (with > 3σ and a B-factor less than 50Å^2).

O was then used to check the water molecules and rebuild the structure. In all three stuctures the alanine residue on the C-termini could not be positioned into density and were therefore removed from the PDBs. Density for the N-terminus alanine in the C183S *Pf*LDH mutant could not be observed and so was also removed.

Whilst rebuilding the 3 mutant structures, it was noticed that some of the residues comprising the active site loop did not properly fit into electron density.

The refinement procedure of minimisation followed by individual B-factor refinement followed by a water molecule search (with a σ value cutoff of 2.5

for the last round) and manual adjustment of the model in *O* was repeated a further two times.

A summary of the C183V, C183A, and C183S refinement processes, along with the R_{cryst}, R_{free}, rmsd bond lengths and angles values at each stage, are shown in Tables 7.13, 7.14 and 7.15, respectively.

It was evident from the R_{crsyst} and R_{free} values that the C183A mutant *Pf*LDH structural data had been 'overfitted' during the final water search and manual rebuild, therefore the PDB coordinates after the preceding individual B-factor refinement were taken as the final model. The C183V model after the penultimate the rebuild so this was taken as the final model. The structure after the penultimate individual B-factor refinement step was taken as the model for the C183S *Pf*LDH mutant.

The final refinement statistics for each of the three mutant structures are shown in Tables 7.16 - 7.18.

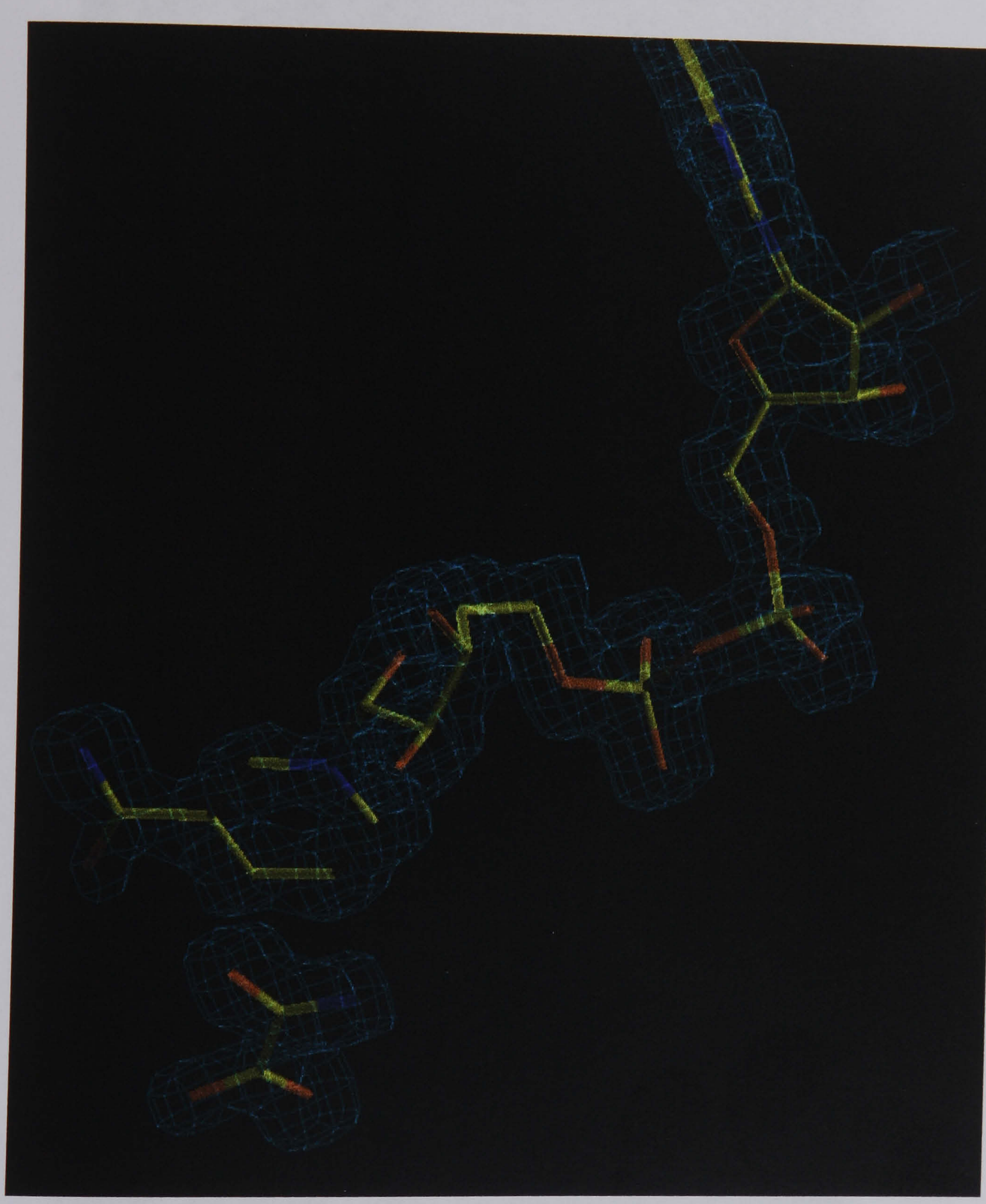
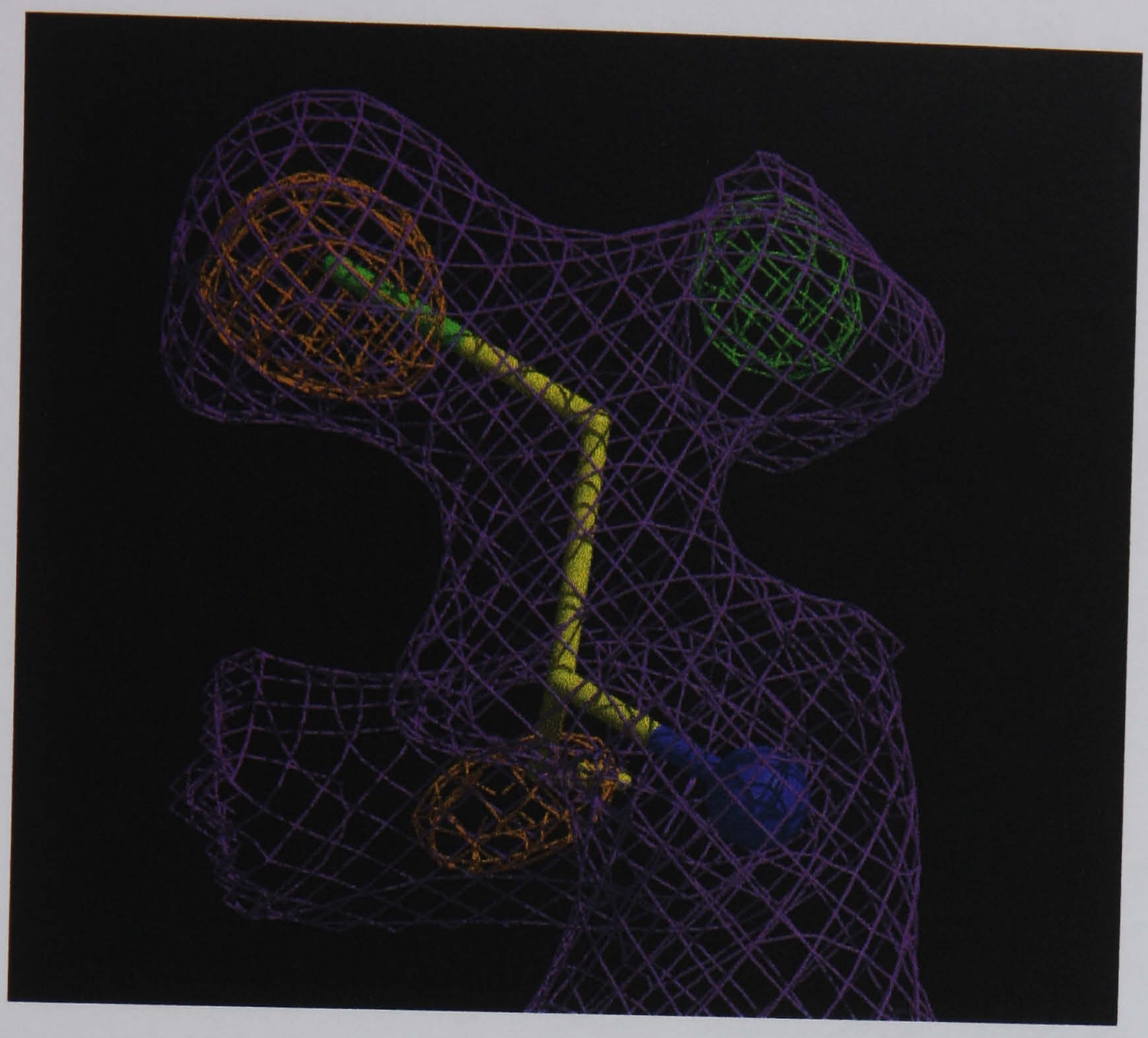


Figure 7.3 *SIGMAA* (CCP4) weighted $2|F_{obs}|-|F_{calc}|$ map (contoured at 1.5σ) with phases from the final C183A *Pf*LDH model. Electron density for NADH and oxamate can clearly be seen in all three C183 mutant models. Image produced using *O*.

(a)



(b)

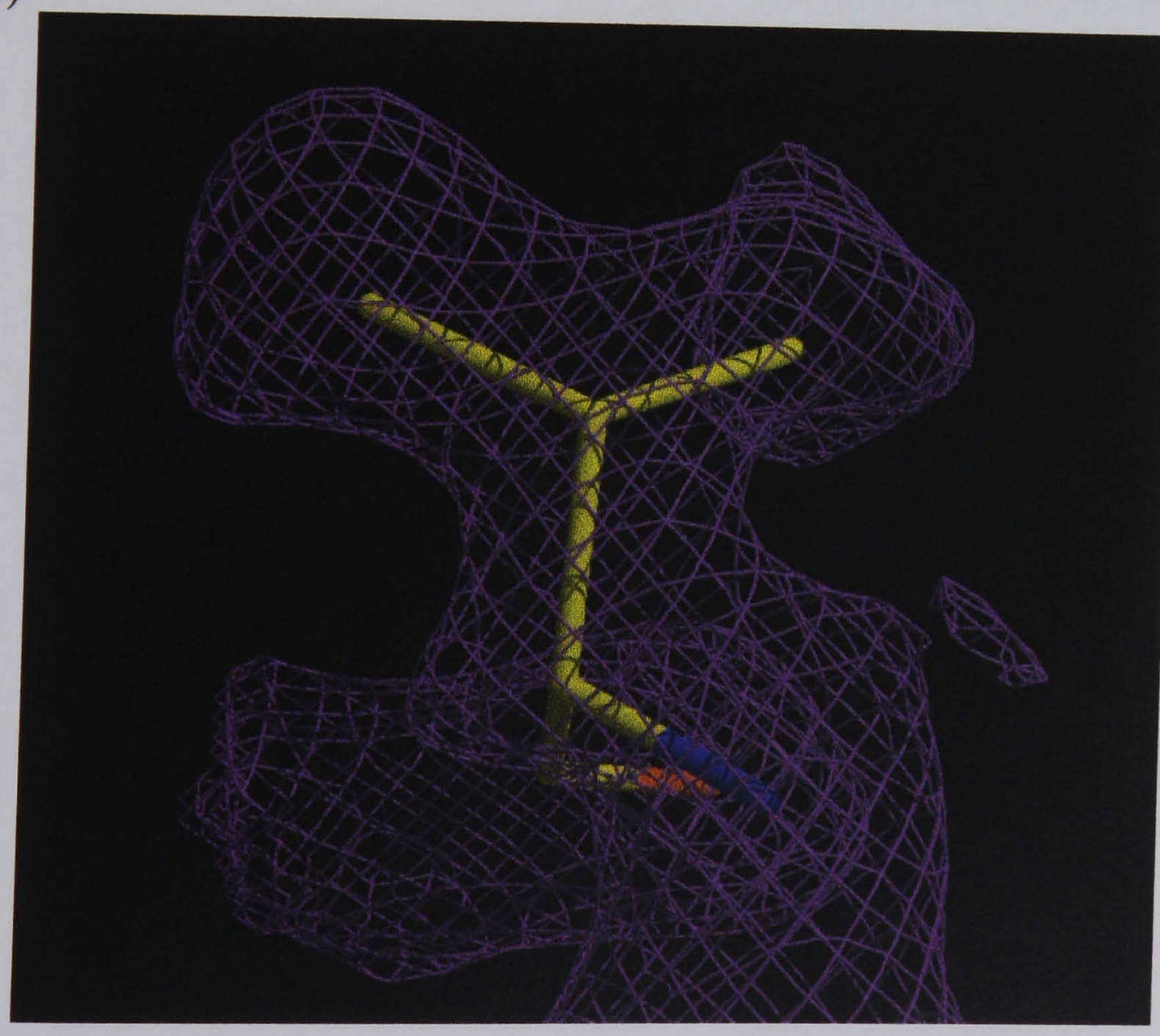
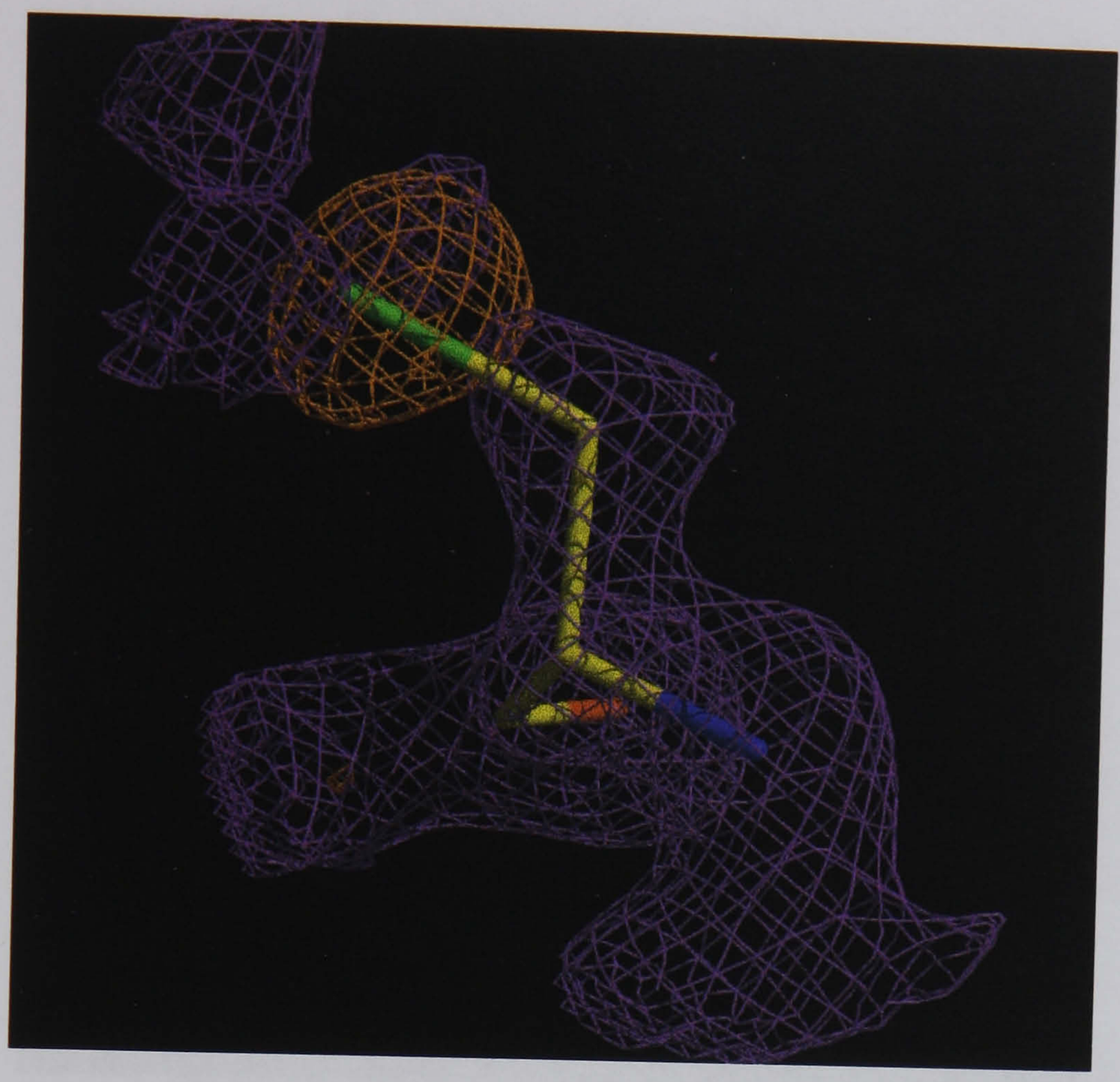


Figure 7.4 (a) *SIGMAA* weighted $2|F_{obs}|-|F_{calc}|$ (purple), $|F_{obs}|-|F_{calc}|$ positive difference (green), and $|F_{obs}|-|F_{calc}|$ negative difference (brown) maps all contoured at 3σ with phases from the final C183V *Pf*LDH model. The positive and negative difference density maps suggest that a valine residue is situated at this locus. (b) Sequencing of the *Pf*LDH gene confirmed that there is a valine at this position. Images produced using the *Xfit* program from the XtalView Suite (McRee, 1999).

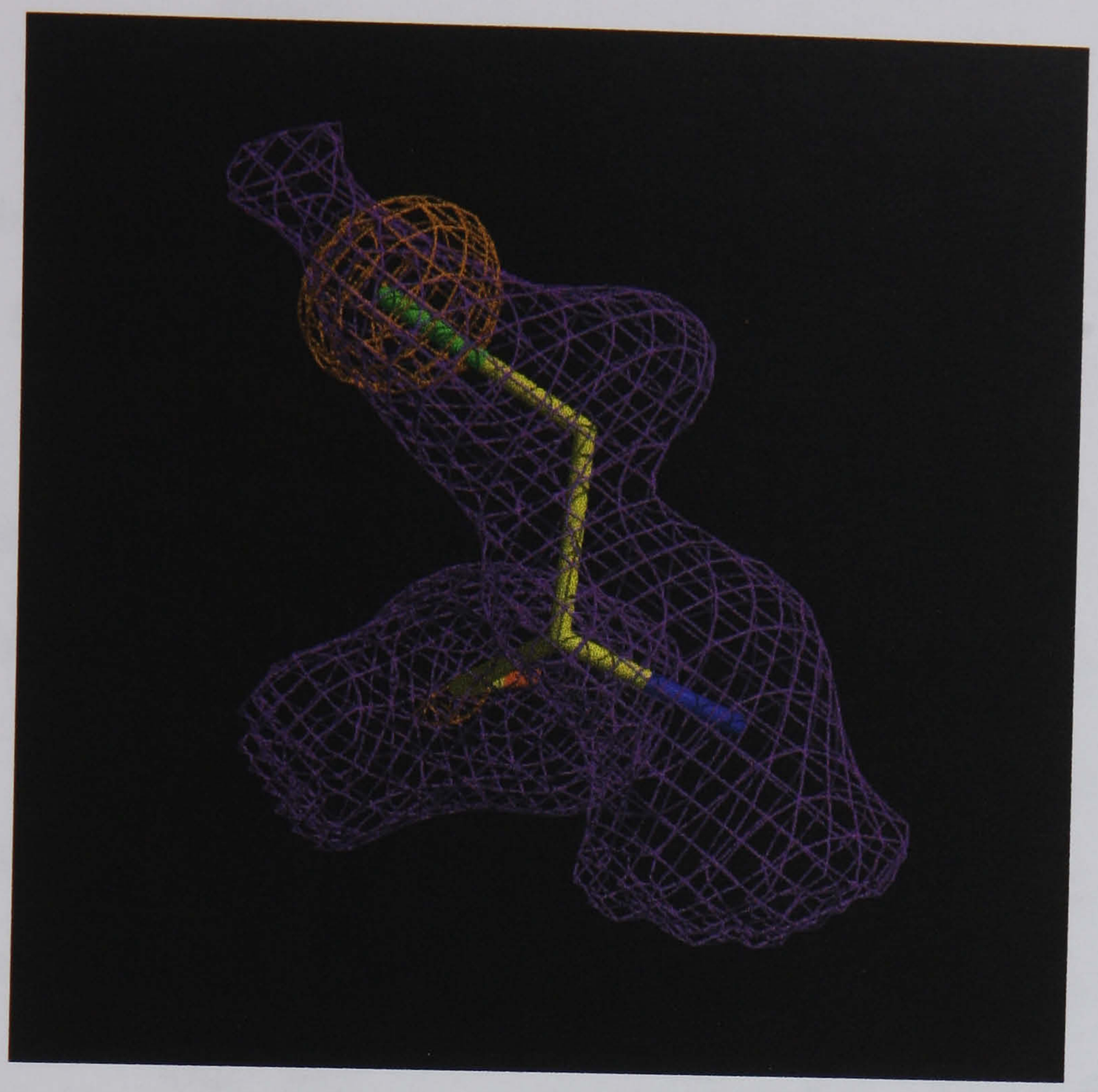
(a)



(b)

Figure 7.5 (a) *SIGMAA* weighted $2|F_{obs}|-|F_{calc}|$ (purple) and $|F_{obs}|-|F_{calc}|$ negative difference (brown) maps both contoured at 3σ with phases from the final C183A *Pf*LDH model. The negative difference density map suggests that a alanine residue is situated at this locus.
(b) Sequencing of the *Pf*LDH gene confirmed that there is a alanine at this position.
Images produced using the *Xfit* program.

(a)



(b)

Figure 7.6 (a) *SIGMAA* weighted $2|F_{obs}|-|F_{calc}|$ (purple) and $|F_{obs}|-|F_{calc}|$ negative difference (brown) maps both contoured at 3σ with phases from the final C183S *Pf*LDH model. The negative difference density map suggests that a serine residue is situated at this locus.
(b) Sequencing of the *Pf*LDH gene confirmed that there is a serine at this position.
Images produced using the *Xfit* program.

Step No.	Refinement Procedure	Final R_{cryst} (%)	Final R_{free} (%)	Geometry bond lengths/ bond angles
1	*AMoRe*- rigid body fiting	29.1	-	-
2	*Refmac*	26.0	29.8	0.013Å/ 1.54°
3a	Simulated annealing- 2500K	27.0	30.3	-
3b	Individual B-factor refinement	25.3	28.9	0.007Å/ 1.27°
4a	Simulated annealing- 2500K	25.1	28.9	-
4b	*water_pick*	23.7	27.5	0.006Å/ 1.21°
5a	Rebuild followed by minimisation	22.3	25.6	-
5b	Individual B-factor refinement	21.9	25.3	-
5c	*water_pick*	21.4	24.8	0.006Å/ 1.20°
6a	Rebuild followed by minimisation	21.5	25.0	-
6b	Individual B-factor refinement	21.3	24.9	-
6c	*water_pick*	20.1	23.6	0.006Å/ 1.17°
7a	Rebuild followed by minimisation	17.7	21.1	-
7b	Individual B-factor refinement	17.5	20.9	-
7c	*water_pick*	17.1	20.8	0.005Å/ 1.14°
8a	Rebuild followed by minimisation	17.0	20.8	-
8b	Individual B-factor refinement	16.9	20.8	-
8c	*water_pick*	16.9	20.6	0.005Å/ 1.13°
9a	Rebuild followed by minimisation	16.9	20.7	-
9b	Individual B-factor refinement	17.0	20.6	-
9c	*water_pick*	17.0	20.6	0.005Å/ 1.13°
10	Rebuild followed by minimisation	17.1	20.8	0.005Å/ 1.14°

Table 7.13 Summary of crystallographic refinement and rebuilding of the C183V mutant.

Step No.	Refinement Procedure	Final R_{cryst} (%)	Final R_{free} (%)	Geometry bond lengths/ bond angles
1	*AMoRe*- rigid body fiting	27.5	-	-
2	*Refmac*	26.9	29.8	0.011Å/ 1.49°
3a	Simulated annealing- 2500K	28.5	31.2	-
3b	Individual B-factor refinement	26.4	29.3	0.006Å/ 1.24°
4a	Simulated annealing- 2500K	26.3	29.4	-
4b	*water_pick*	26.1	29.3	0.005Å/ 1.20°
5a	Rebuild followed by minimisation	26.1	29.3	-
5b	Individual B-factor refinement	26.1	29.4	-
5c	*water_pick*	24.3	27.7	0.005Å/ 1.?8°
6a	Rebuild followed by minimisation	23.3	26.5	-
6b	Individual B-factor refinement	23.2	26.4	-
6c	*water_pick*	21.3	24.0	0.005Å/ 1.18°
7a	Rebuild followed by minimisation	21.2	23.7	-
7b	Individual B-factor refinement	21.0	23.5	-
7c	*water_pick*	20.1	22.4	0.005Å/ 1.16°
8a	Rebuild followed by minimisation	18.6	21.2	-
8b	Individual B-factor refinement	18.4	21.0	-
8c	*water_pick*	18.0	20.7	0.005Å/ 1.18°
9a	Rebuild followed by minimisation	17.8	20.7	-
9b	Individual B-factor refinement	17.7	20.5	-
9c	*water_pick*	17.7	20.4	0.005Å/ 1.16°
10	Rebuild followed by minimisation	17.7	20.4	0.005Å/ 1.15°

Table 7.14 Summary of crystallographic refinement and rebuilding of the C183A mutant.

Step No.	Refinement Procedure	Final R_{cryst} (%)	Final R_{free} (%)	Geometry bond lengths/ bond angles
1	*AMoRe*- rigid body fiting	27.5	-	-
2	*Refmac*	25.3	31.8	0.010Å/ 1.47°
3a	Simulated annealing- 2500K	27.0	32.6	-
3b	Individual B-factor refinement	25.4	30.7	0.007Å/ 1.21°
4a	Simulated annealing- 2500K	25.1	30.6	-
4b	*water_pick*	23.9	29.3	0.006Å/ 1.20°
5a	Rebuild followed by minimisation	21.4	26.0	-
5b	Individual B-factor refinement	21.0	25.7	-
5c	*water_pick*	19.2	24.2	0.006Å/ 1.18°
6a	Rebuild followed by minimisation	18.9	24.0	-
6b	Individual B-factor refinement	18.7	24.0	-
6c	*water_pick*	18.0	23.2	0.006Å/ 1.14°
7a	Rebuild followed by minimisation	17.8	22.8	-
7b	Individual B-factor refinement	17.7	22.9	-
7c	*water_pick*	17.5	22.8	0.005Å/ 1.12°
8a	Rebuild followed by minimisation	17.5	22.7	-
8b	Individual B-factor refinement	17.5	22.7	-
8c	*water_pick*	17.5	22.7	0.005Å/ 1.13°
9a	Rebuild followed by minimisation	17.5	22.6	-
9b	Individual B-factor refinement	17.5	22.6	-
9c	*water_pick*	17.6	22.7	0.005Å/ 1.12°
10	Rebuild followed by minimisation	17.6	22.7	0.005Å/ 1.12°

Table 7.15 Summary of crystallographic refinement and rebuilding of the C183S mutant.

Resolution range	30.0 - 2.1Å
R_{cryst}	16.9%
R_{free}	20.7%
rmsd bond lengths	0.005Å
rmsd bond angles	1.13°
Total number of non-H protein atoms	2386
Total number of NADH atoms	44
Total number of oxamate atoms	6
Total number of glycerol atoms	6
Identified number of water molecules	242
Average main-chain/side-chain B-factor	11.65Å^2 / 13.00Å^2
Average NADH B-factor	11.70Å^2
Average oxamate B-factor	14.56Å^2
Average glycerol B-factor	10.87Å^2
Average B-factor for water molecules	21.10Å^2

Table 7.16 Final refinement statistics for the C183V *Pf*LDH mutant.

Resolution range	22.0 - 1.7Å
R_{cryst}	17.7%
R_{free}	20.5%
rmsd bond lengths	0.005Å
rmsd bond angles	1.16°
Total number of non-H protein atoms	2379
Total number of NADH atoms	44
Total number of oxamate atoms	6
Total number of glycerol atoms	6
Identified number of water molecules	324
Average main-chain/side-chain B-factor	6.87Å^2 / 8.96Å^2
Average NADH B-factor	8.23Å^2
Average oxamate B-factor	9.01Å^2
Average glycerol B-factor	13.69Å^2
Average B-factor for water molecules	16.94Å^2

Table 7.17 Final refinement statistics for the C183A *Pf*LDH mutant.

Resolution range	22.0 - 1.9Å
R_{cryst}	17.5%
R_{free}	22.7%
rmsd bond lengths	0.005Å
rmsd bond angles	1.13°
Total number of non-H protein atoms	2385
Total number of NADH atoms	44
Total number of oxamate atoms	6
Total number of glycerol atoms	6
Identified number of water molecules	273
Average main-chain/side-chain B-factor	5.84Å^2 / 7.14Å^2
Average NADH B-factor	11.51Å^2
Average oxamate B-factor	14.25Å^2
Average glycerol B-factor	6.49Å^2
Average B-factor for water molecules	16.71Å^2

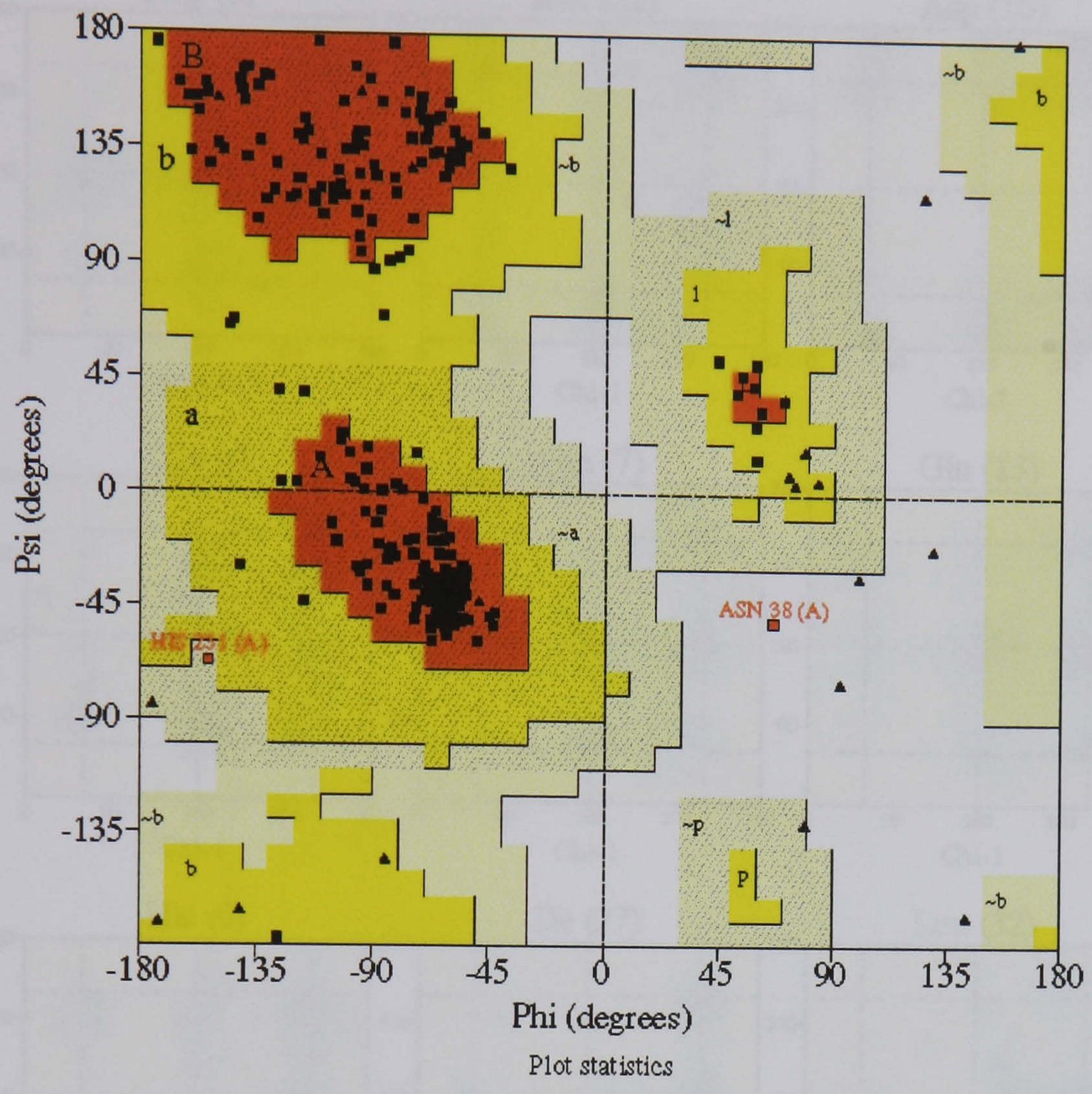
Table 7.18 Final refinement statistics for the C183S *Pf*LDH mutant.

7.6 Validation of the C183 Mutant *Pf*LDH Models

The quality of the final C183 mutant *Pf*LDH models were assessed using the CCP4 suite of programs *PROCHECK* (Laskowski *et al.*, 1993) (Figures 7.7 - 7.15). *PROCHECK* was also used during refinement to evaluate the current model.

The Ramachandran plot (Ramachandran and Sasisekharen, 1968) shows that in all three models 90% of all the non-proline residues are in the favoured regions and the overwhelming majority of the remaining regions are in the allowed regions. The Chi1-Chi2 plots indicate that the models conform well to standard geometric criteria for their particular resolution ranges. Additionally, statistics for the main chain and side chain parameters are also shown.

(a)



Plot statistics

Residues in most favoured regions [A,B,L]	252	91.6%
Residues in additional allowed regions [a,b,l,p]	21	7.6%
Residues in generously allowed regions [~a,~b,~l,~p]	1	0.4%
Residues in disallowed regions	1	0.4%
Number of non-glycine and non-proline residues	275	100.0%
Number of end-residues (excl. Gly and Pro)	4	
Number of glycine residues (shown as triangles)	26	
Number of proline residues	12	
Total number of residues	317	

(b)

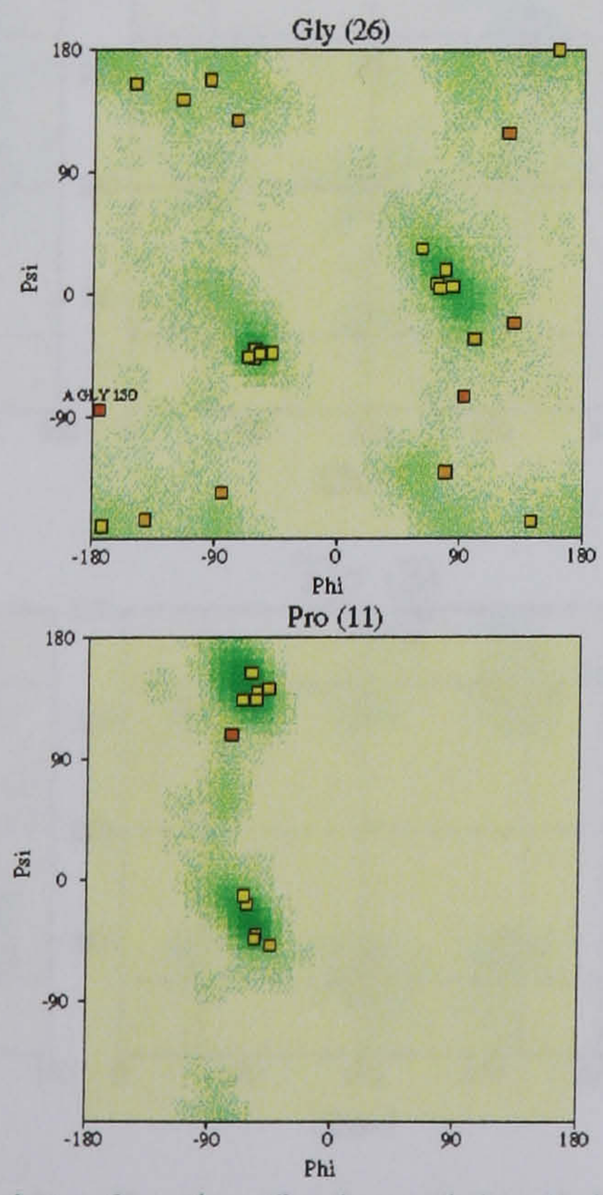


Figure 7.7 Ramachandran plots for the final model of the C183V *Pf*LDH mutant, as produced by *PROCHECK*. (a) Ramachandran plot for non proline residues, (glycines shown as triangles). (b) Ramachandran plot for glycine and proline residues.

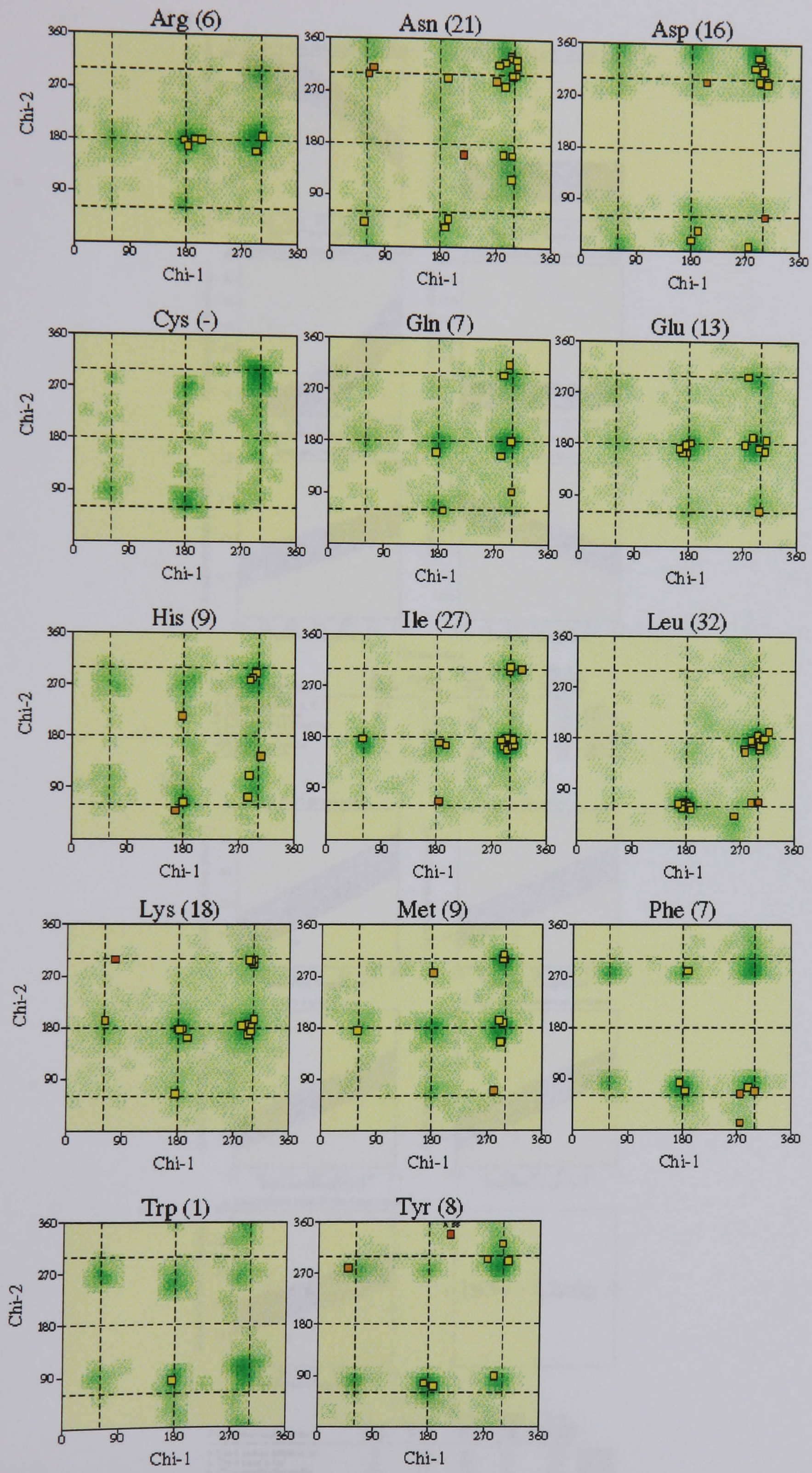
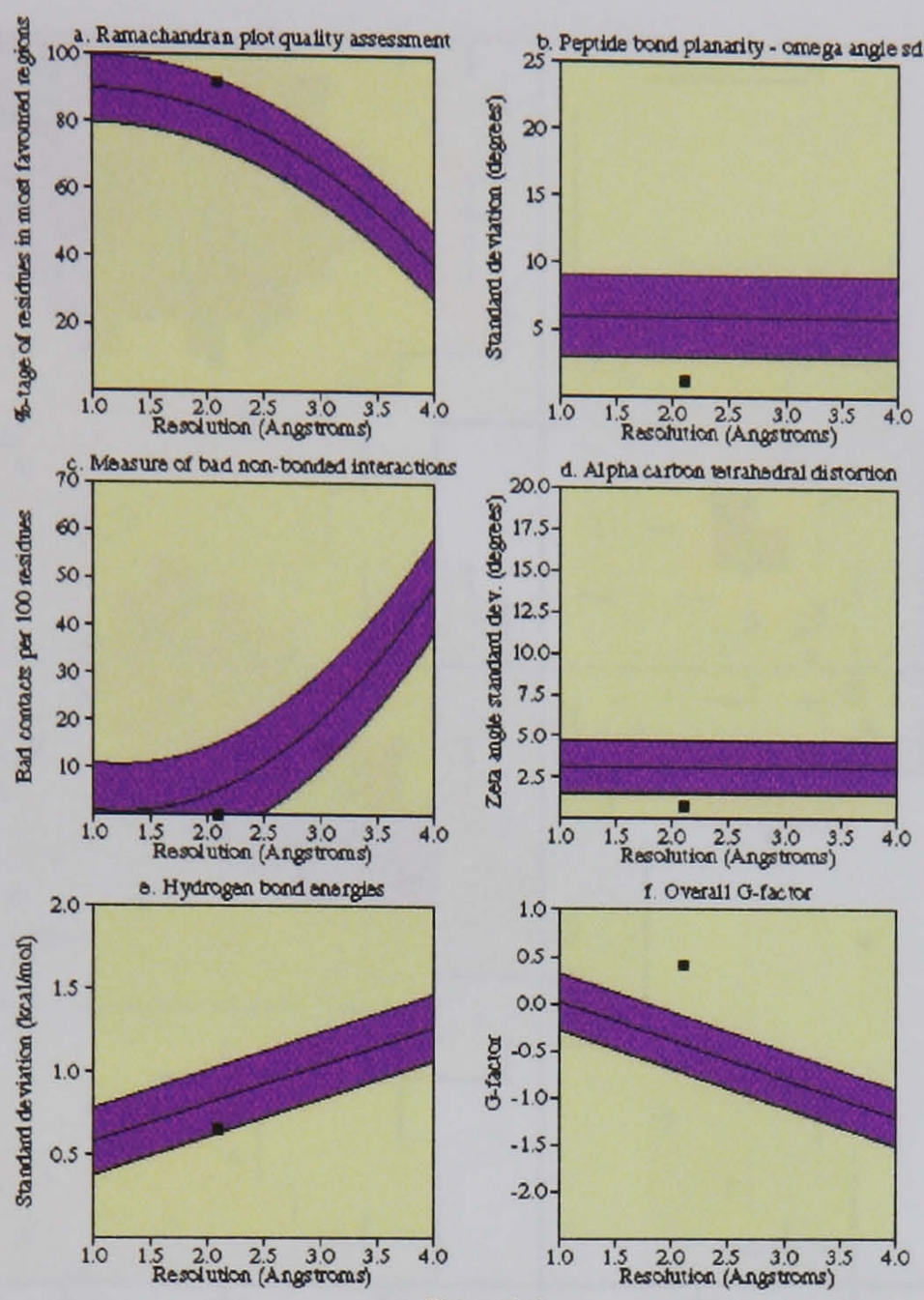


Figure 7.8 Chi1 versus Chi2 plots for the residues in the final model of the C183V *Pf*LDH mutant, produced with *PROCHECK*.

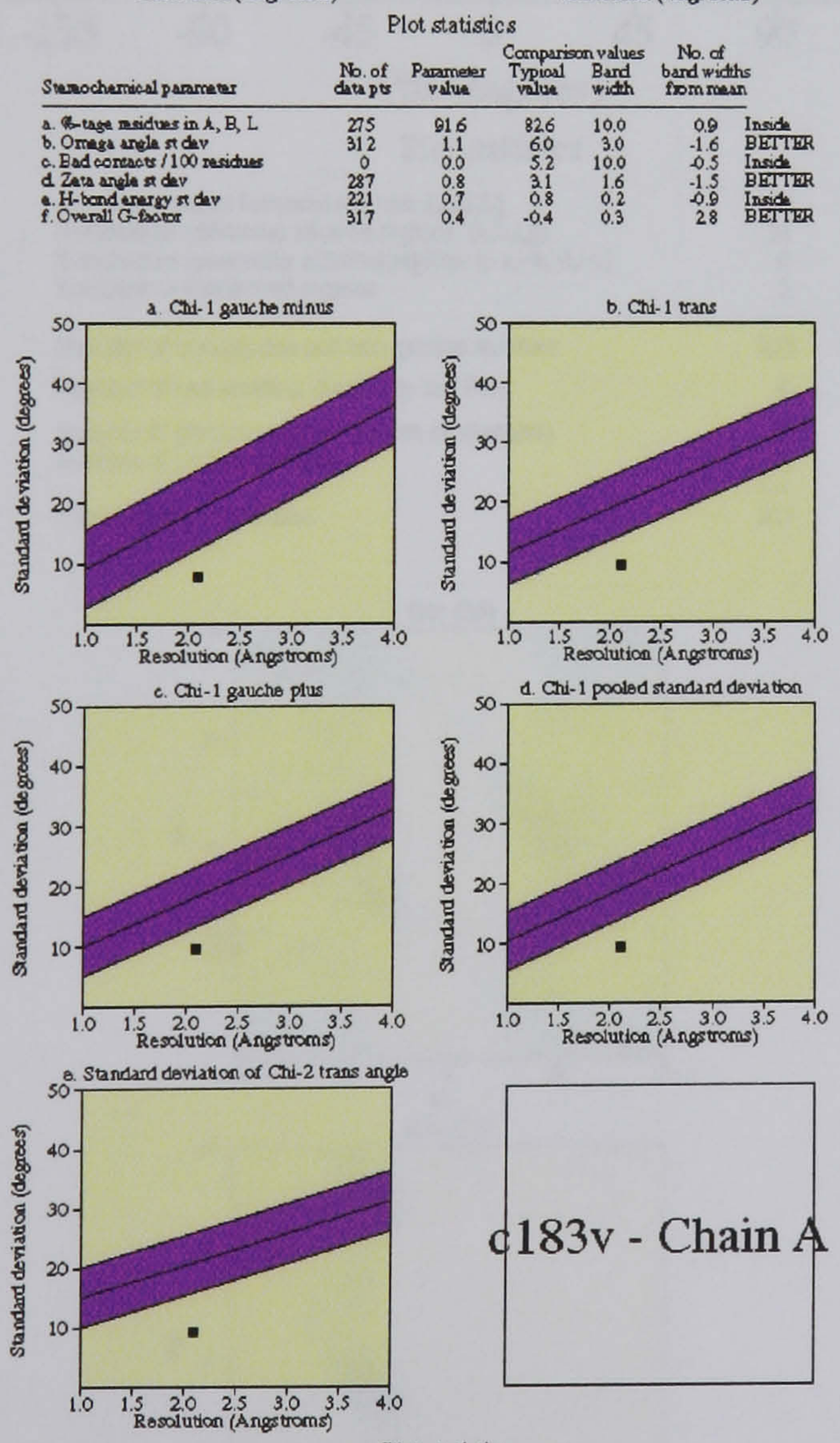
(a)



Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values Typical value	Band width	No. of band widths from mean	
a. %-tage residues in A, B, L	275	91.6	82.6	10.0	0.9	Inside
b. Omega angle st dev	312	1.1	6.0	3.0	-1.6	BETTER
c. Bad contacts / 100 residues	0	0.0	5.2	10.0	-0.5	Inside
d. Zeta angle st dev	287	0.8	3.1	1.6	-1.5	BETTER
e. H-bond energy st dev	221	0.7	0.8	0.2	-0.9	Inside
f. Overall G-factor	317	0.4	-0.4	0.3	2.8	BETTER

(b)

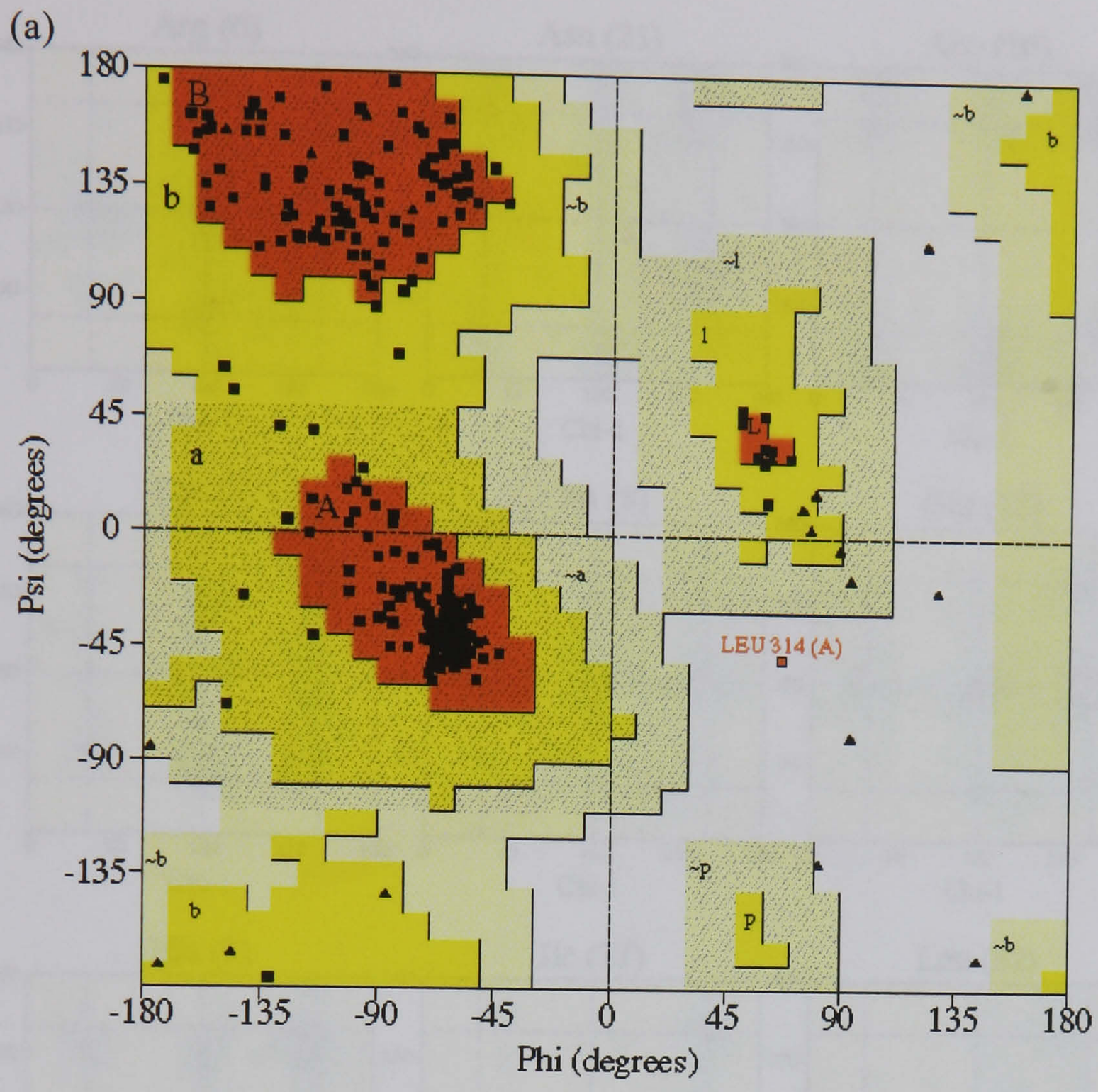


Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values Typical value	Band width	No. of band widths from mean	
a. Chi-1 gauche minus st dev	25	7.9	19.0	6.5	-1.7	BETTER
b. Chi-1 trans st dev	84	9.4	19.7	5.3	-2.0	BETTER
c. Chi-1 gauche plus st dev	137	9.5	18.3	4.9	-1.8	BETTER
d. Chi-1 pooled st dev	246	9.5	18.9	4.8	-2.0	BETTER
e. Chi-2 trans st dev	77	9.2	20.9	5.0	-2.3	BETTER

Figure 7.9 (a) Main chain and (b) side chain statistics for the final model of the C183V *Pf*LDH mutant, produced with *PROCHECK*.

(a)



Plot statistics

Residues in most favoured regions [A,B,L]	251	91.9%
Residues in additional allowed regions [a,b,l,p]	21	7.7%
Residues in generously allowed regions [~a,~b,~l,~p]	0	0.0%
Residues in disallowed regions	1	0.4%
Number of non-glycine and non-proline residues	273	100.0%
Number of end-residues (excl. Gly and Pro)	6	
Number of glycine residues (shown as triangles)	26	
Number of proline residues	12	
Total number of residues	317	

(b)

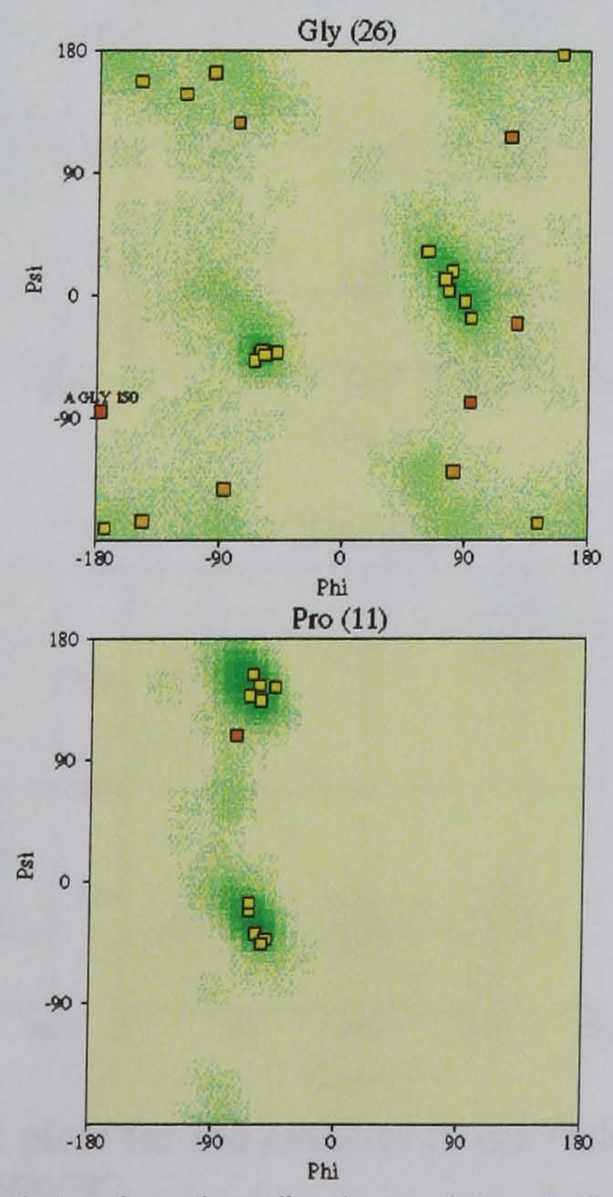


Figure 7.10 Ramachandran plots for the final model of the C183A *Pf*LDH mutant, as produced by *PROCHECK*. (a) Ramachandran plot for non proline residues, (glycines shown as triangles). (b) Ramachandran plot for glycine and proline residues.

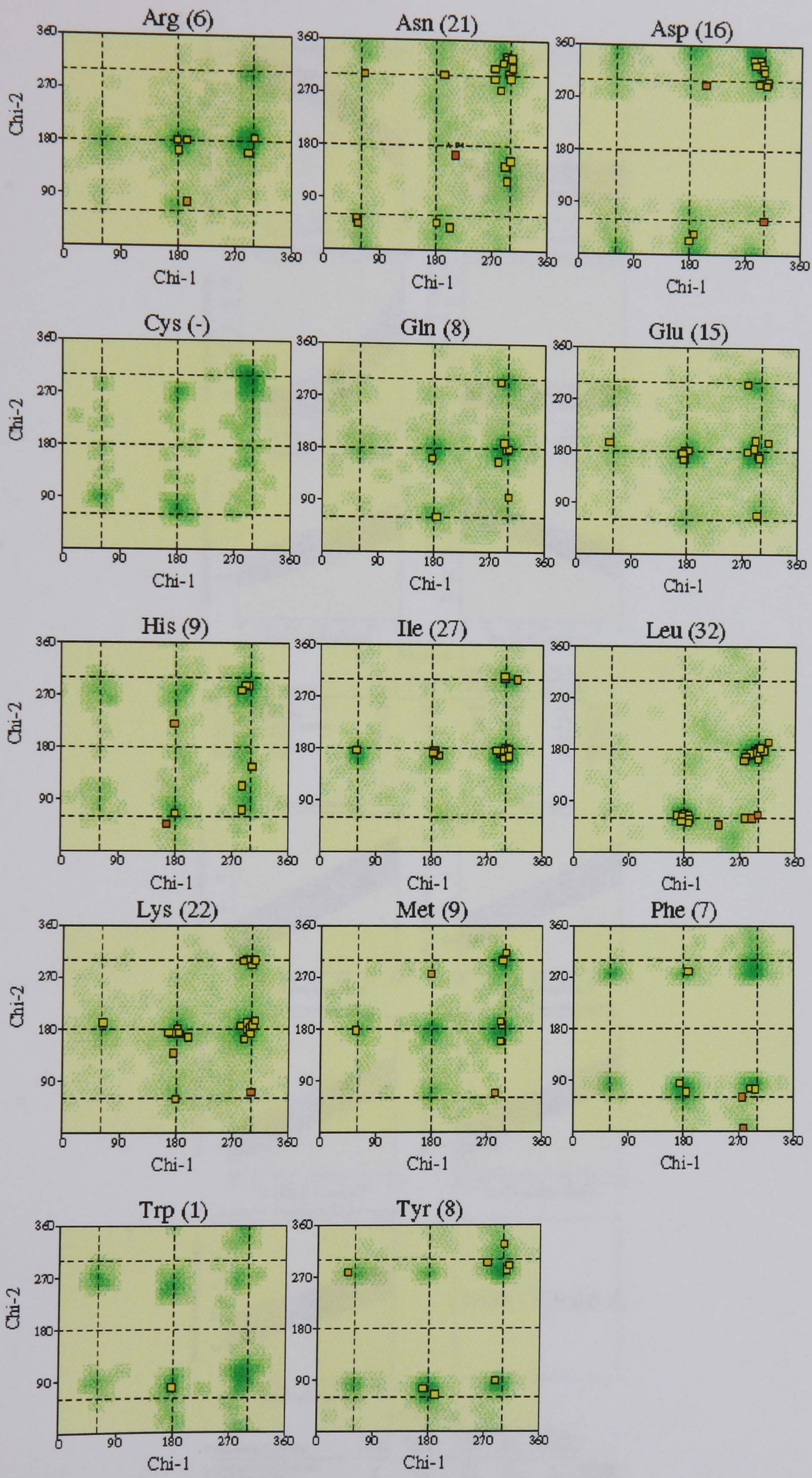
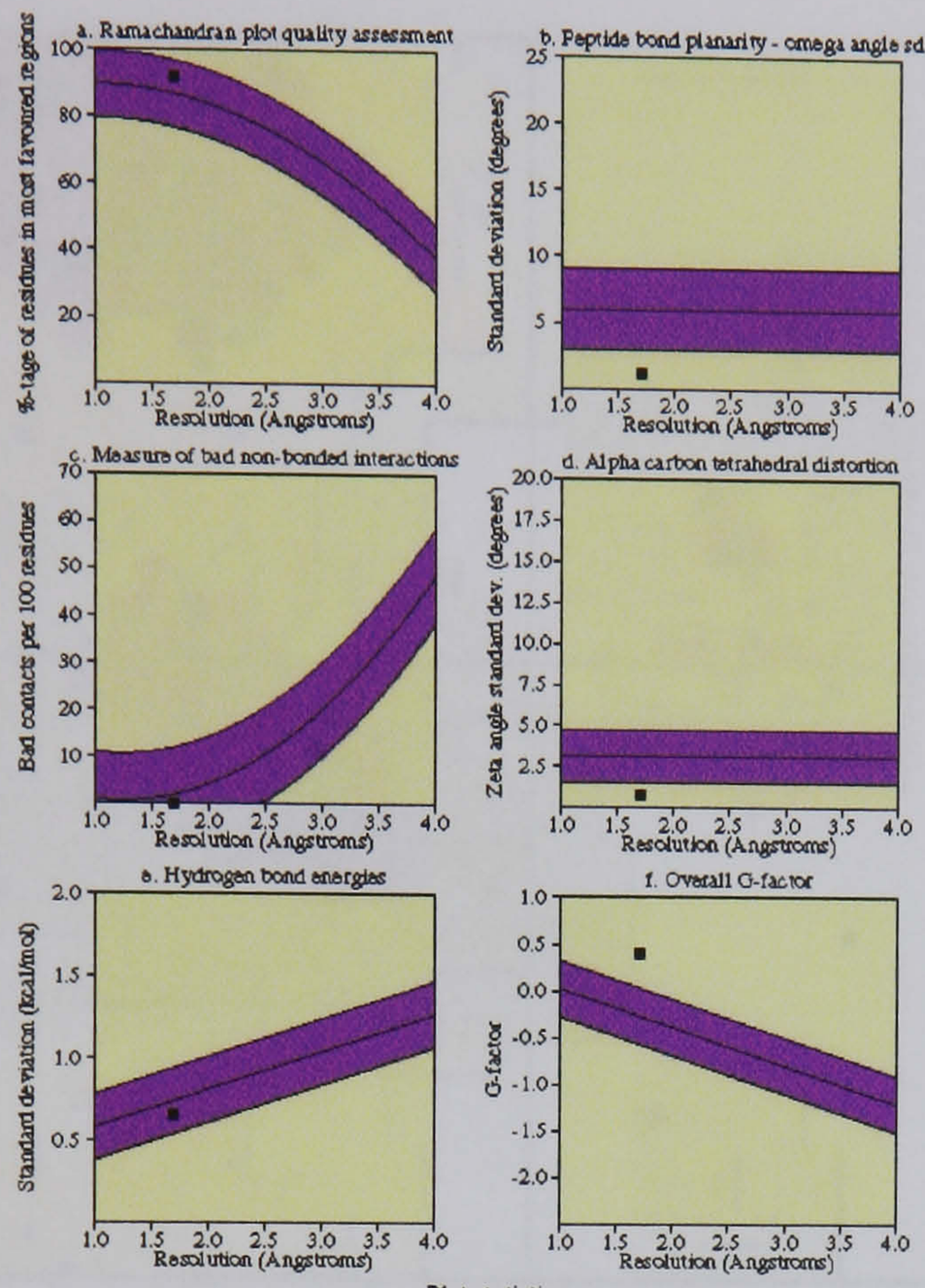


Figure 7.11 Chi1 versus Chi2 plots for the residues in the final model of the C183A *Pf*LDH mutant, produced with *PROCHECK*.

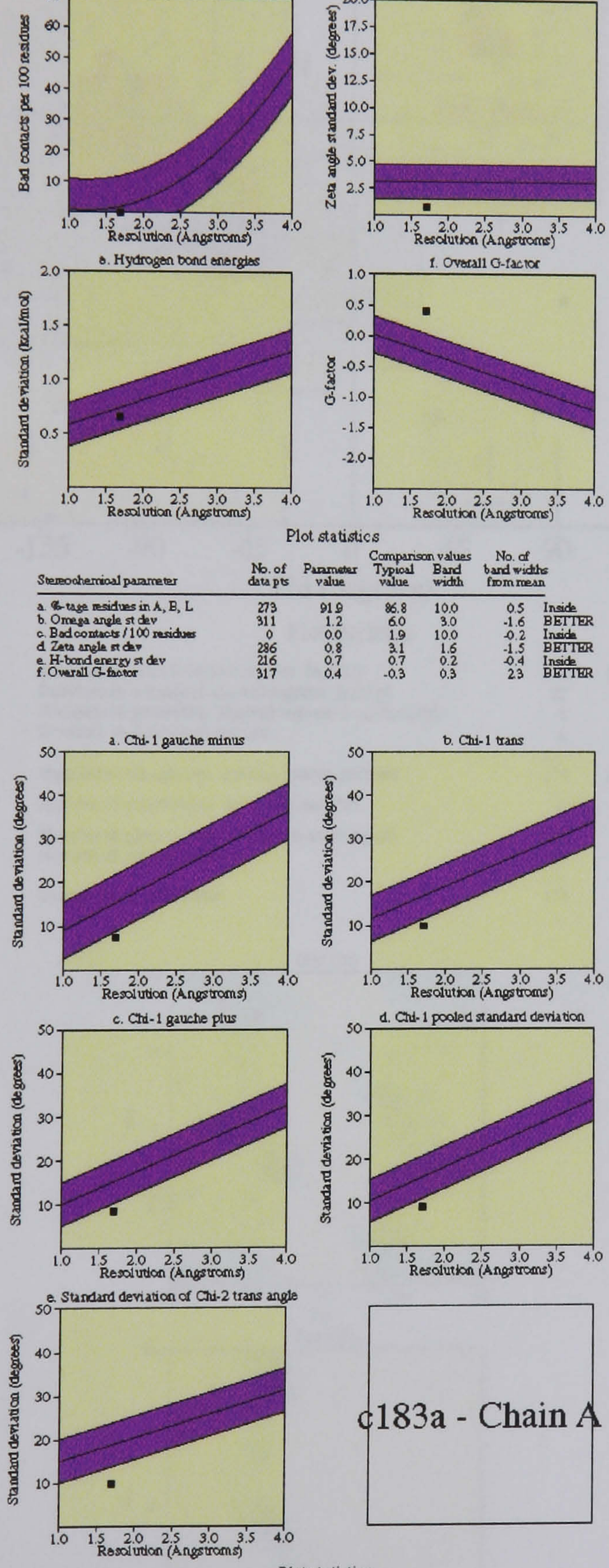
(a)



Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values Typical value	Band width	No. of band widths from mean	
a. %-tage residues in A, B, L	273	91.9	86.8	10.0	0.5	Inside
b. Omega angle st dev	311	1.2	6.0	3.0	-1.6	BETTER
c. Bad contacts / 100 residues	0	0.0	1.9	10.0	-0.2	Inside
d. Zeta angle st dev	286	0.8	3.1	1.6	-1.5	BETTER
e. H-bond energy st dev	216	0.7	0.7	0.2	-0.4	Inside
f. Overall G-factor	317	0.4	-0.3	0.3	2.3	BETTER

(b)

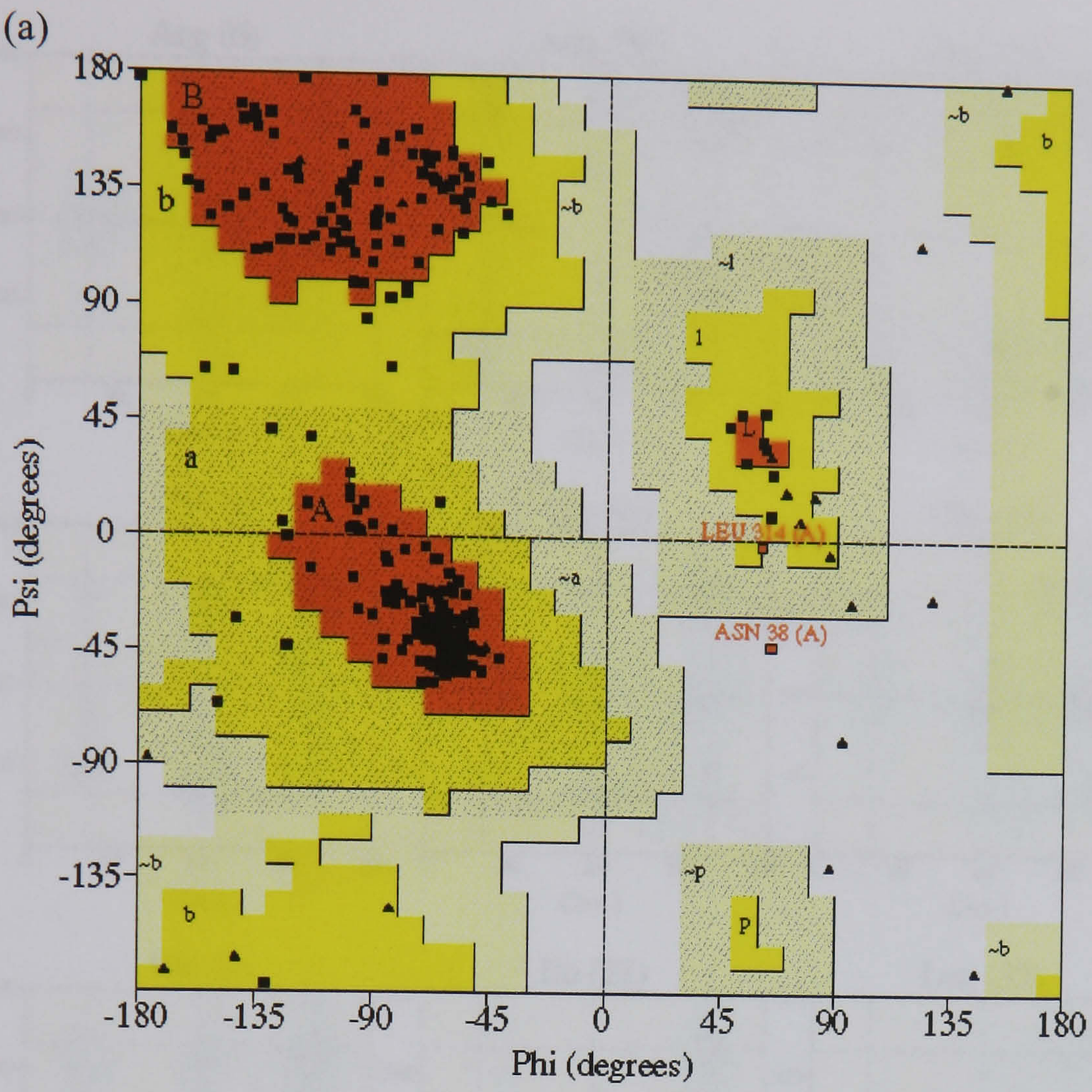


Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values Typical value	Band width	No. of band widths from mean	
a. Chi-1 gauche minus st dev	24	7.4	15.4	6.5	-1.2	BETTER
b. Chi-1 trans st dev	86	9.9	16.8	5.3	-1.3	BETTER
c. Chi-1 gauche plus st dev	136	8.5	15.3	4.9	-1.4	BETTER
d. Chi-1 pooled st dev	246	9.1	15.9	4.8	-1.4	BETTER
e. Chi-2 trans st dev	84	9.9	18.8	5.0	-1.8	BETTER

Figure 7.12 (a) Main chain and (b) side chain statistics for the final model of the C183A *Pf*LDH mutant, produced with *PROCHECK*.

(a)



Plot statistics

Residues in most favoured regions [A,B,L]	248	90.2%
Residues in additional allowed regions [a,b,l,p]	25	9.1%
Residues in generously allowed regions [~a,~b,~l,~p]	1	0.4%
Residues in disallowed regions	1	0.4%
Number of non-glycine and non-proline residues	275	100.0%
Number of end-residues (excl. Gly and Pro)	4	
Number of glycine residues (shown as triangles)	26	
Number of proline residues	12	
Total number of residues	317	

(b)

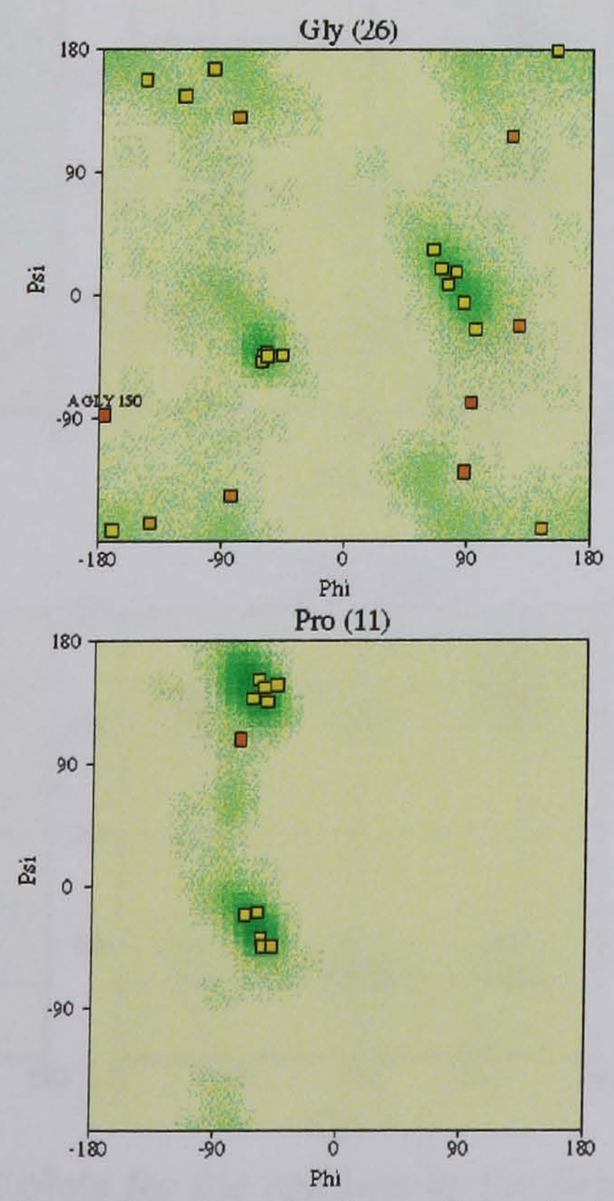


Figure 7.13 Ramachandran plots for the final model of the C183S *Pf*LDH mutant, as produced by *PROCHECK*. (a) Ramachandran plot for non proline residues, (glycines shown as triangles). (b) Ramachandran plot for glycine and proline residues.

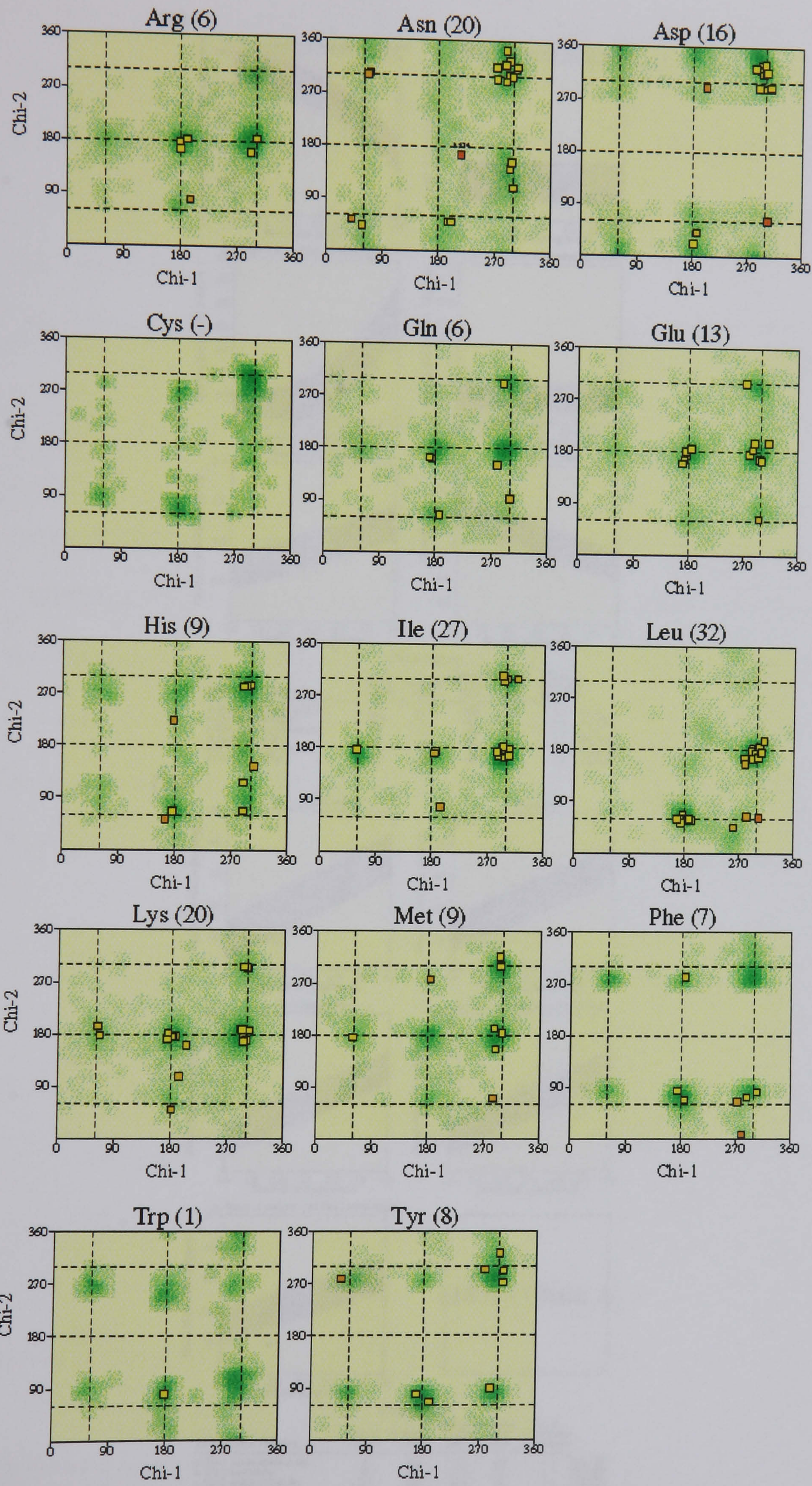
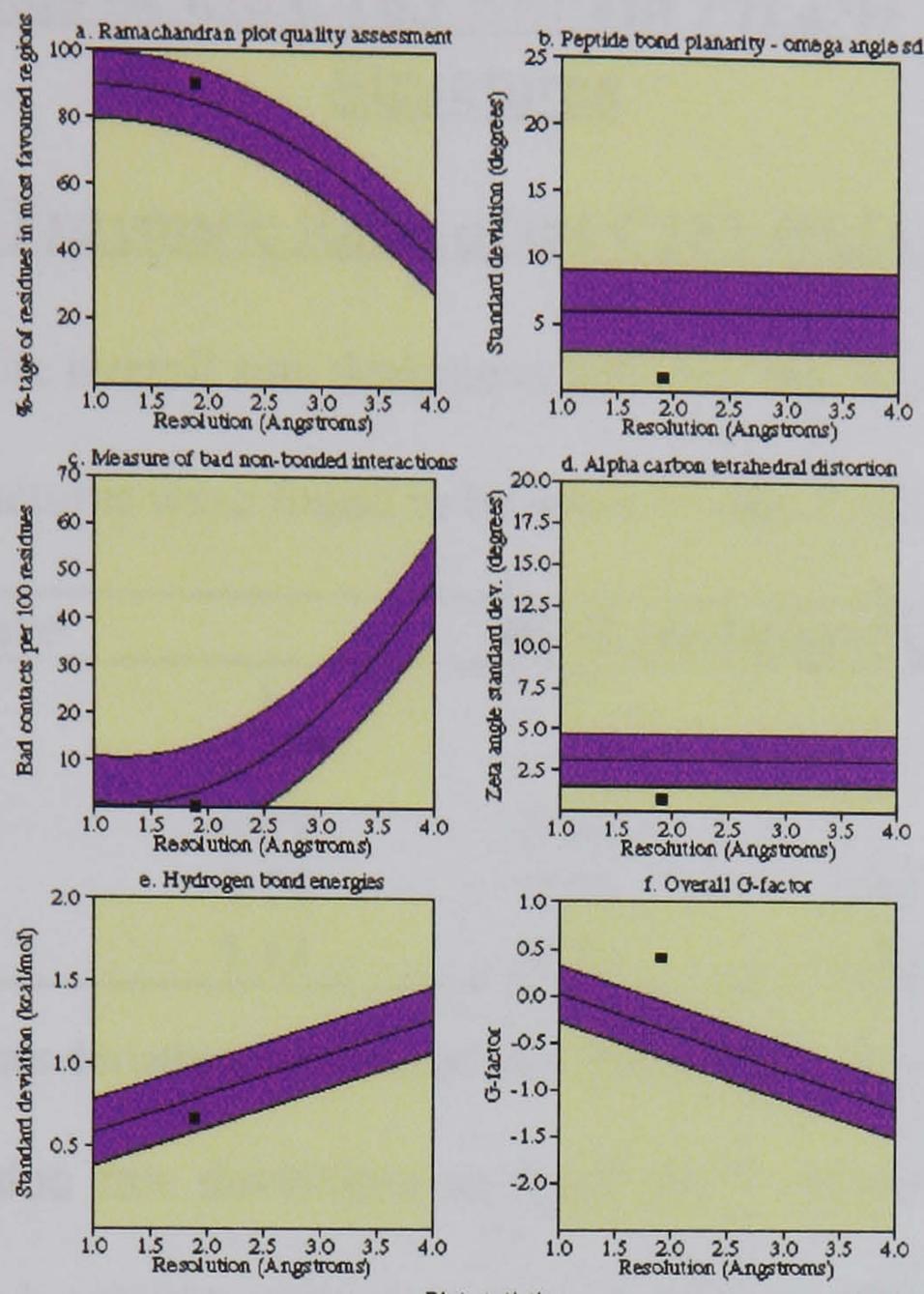


Figure 7.14 Chi1 versus Chi2 plots for the residues in the final model of the C183S *Pf*LDH mutant, produced with *PROCHECK*.

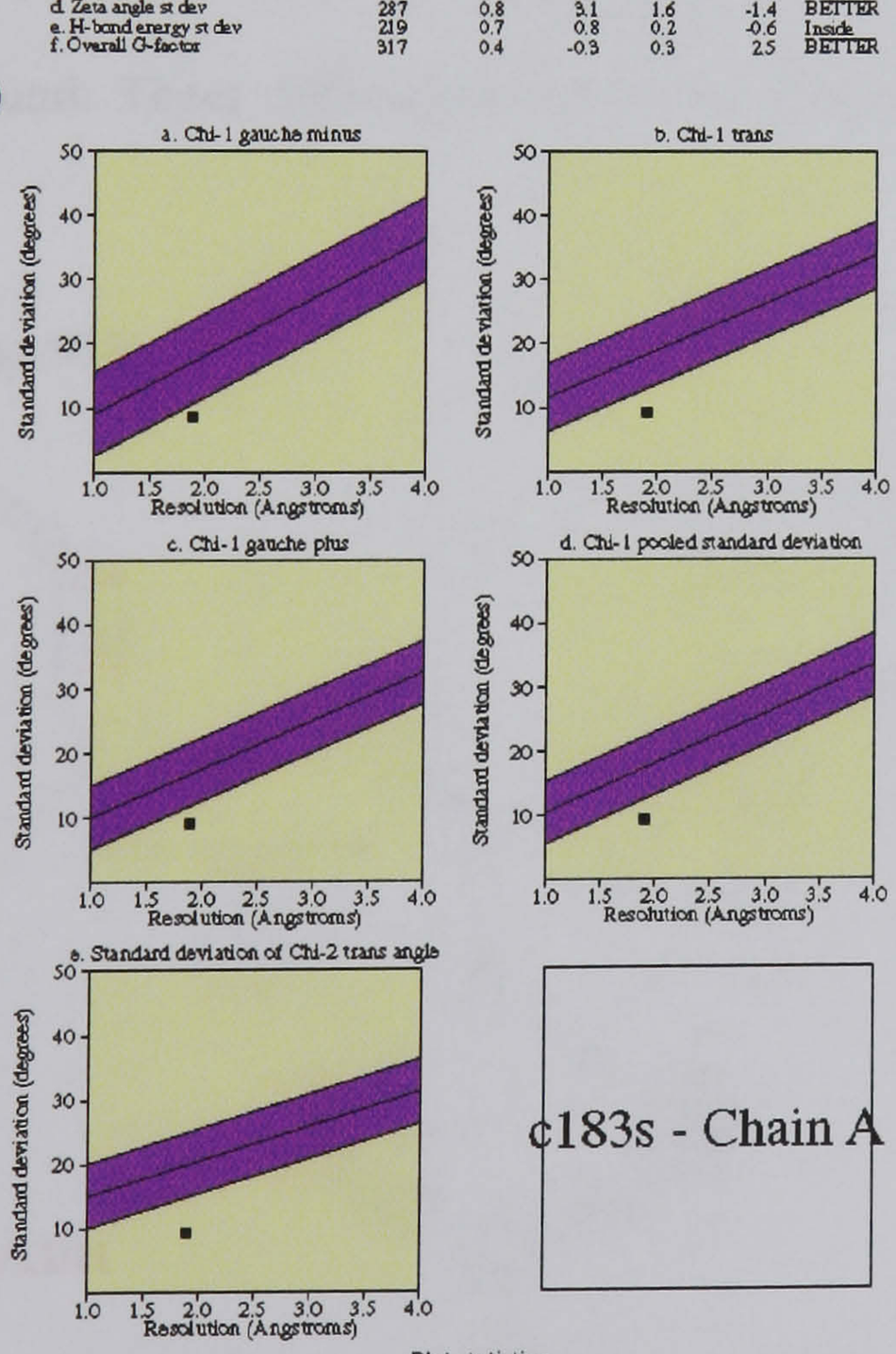
(a)



Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values Typical value	Band width	No. of band widths from mean	
a. %-tage residues in A, B, L	275	90.2	84.9	10.0	0.5	Inside
b. Omega angle st dev	312	1.1	6.0	3.0	-1.6	BETTER
c. Bad contacts / 100 residues	1	0.3	3.3	10.0	-0.3	Inside
d. Zeta angle st dev	287	0.8	3.1	1.6	-1.4	BETTER
e. H-bond energy st dev	219	0.7	0.8	0.2	-0.6	Inside
f. Overall G-factor	317	0.4	-0.3	0.3	2.5	BETTER

(b)



Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values Typical value	Band width	No. of band widths from mean	
a. Chi-1 gauche minus st dev	26	8.6	17.2	6.5	-1.3	BETTER
b. Chi-1 trans st dev	84	9.2	18.3	5.3	-1.7	BETTER
c. Chi-1 gauche plus st dev	135	9.0	16.8	4.9	-1.6	BETTER
d. Chi-1 pooled st dev	245	9.3	17.4	4.8	-1.7	BETTER
e. Chi-2 trans st dev	78	9.3	19.9	5.0	-2.1	BETTER

Figure 7.15 (a) Main chain and (b) side chain statistics for the final model of the C183S *Pf*LDH mutant, produced with *PROCHECK*.

7.7 Analysis of the C183 Mutant *Pf*LDH Molecular Structures

7.7.1 Reduced Enzymatic Rates of the C183 *Pf*LDH Mutants

As expected, The overall rms deviations between the WT *Pf*LDH and three C183 mutant structures were found to be small (Table 7.19).

***Pf*LDH Structure**	**RMS Deviation (Å)**			
	1	2	3	4
1. WT	0.00			
2. C183V	0.99	0.00		
3. C183A	1.14	0.55	0.00	
4. C183S	1.13	0.51	0.40	0.00

Table 7.19 Overall rms deviations between the wild type and C183 mutant *Pf*LDH structures.

Individual residue rms deviations amongst the 4 crystal structures where then calculated using the Uppsala suite *lsqman* program. Small but significant deviations in the positions of the residues involved in catalysis and substrate binding were found: These differences were then inspected in *O* (Figure 7.16).

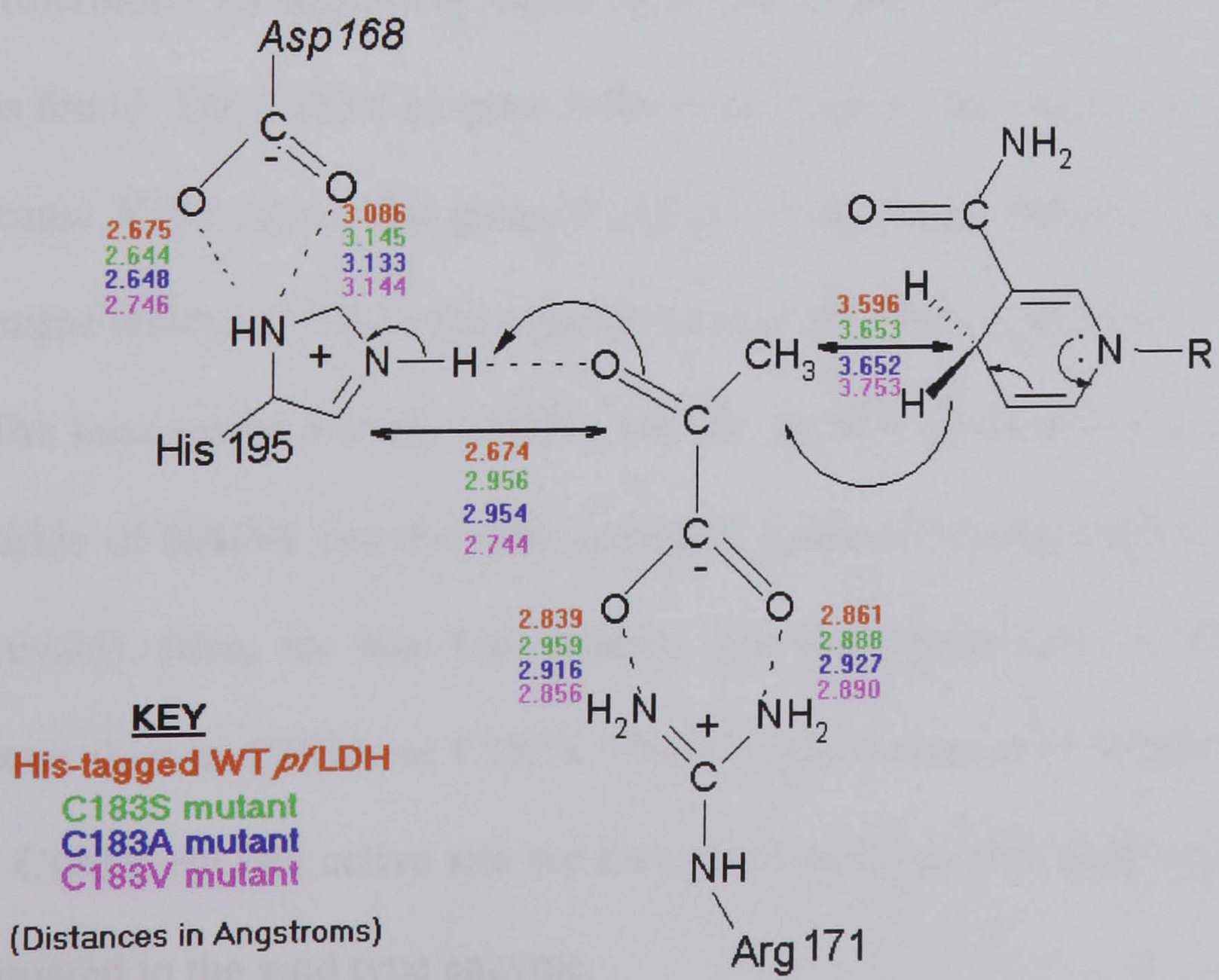


Figure 7.16 Differences in the active sites of the C183 *Pf*LDH mutants. (pyruvate shown in active site instead of the actual oxamate).

Mutation of the cysteine 183 residue causes a decrease in the enzymatic activity of *Pf*LDH: The C183S mutant has the greatest initial activity, followed by the C183A and C183V mutant *Pf*LDH enzymes, respectively (Figure 6.3).

It is unclear how a mutation which is remote in both primary and tertiary structure from the active site can have such as an affect on kinetics.

It seems as though the enzymatic rate of *Pf*LDH is dependent on the degree to which the dimer subunit has been perturbed. The serine mutant has the greatest enzyme activity of the three C183 mutants presumably because it effects the subunit interface architecture the least (the serine residue has an -OH group instead of an -SH group of the cysteine residue). Indeed, two water molecules found closely associated with the C183 residue in the wild type enzyme were found in exactly the same position next to the serine residue in the C183S mutant. In the the C183A mutant the alanine residue is small and so would not alter the structure of the subunit interface. Furthermore, an additional water molecule in the C183A crystal structure was found. The C183V enzyme is the least active of the mutants presumably because V183 causes the greatest change at the dimer subunit interface: A tyrosine residue (73A) in the adjacent subunit has altered position.

The least active mutant, C183V, has the greatest distance between the C4 hydride of NADH and the -NH_2 group of oxamate (or the methyl group of pyruvate). Also, the His-Asp catalytic pair are further apart in the valine mutant than in C183A or C183S. These slight increases in distance within the C183V *Pf*LDH active site are enough to cause a 40% drop in k_{cat} when compared to the wild type enzyme.

The initial activities of the C183A and C183S enzymes are approximately the same. This is probably due to the enzymes having very similar distances between their active site residues and the substrate and cofactor molecules.

7.7.2 Oxidative Stability of the C183 *Pf*LDH Mutants

The rationale for mutating the cysteine 183 residue in *Pf*LDH was three-fold (Richard Sessions, personal communication). Firstly, of the four cysteines in the enzyme only C76 and C183 where at the dimer subunit interface, therefore these residues were the most likely to form an inappropriate inter-subunit disulphide bond. However, the C183 and C76 residues are too far apart (about 9.5Å) in order to form a direct disulphide bond (Figures 7.17 and 7.18).(The C183 residues are about 20Å apart). It was presumed that the β-sheet on which C76 sits could undergo a minor conformational bringing the residue into apposition with C183. Finally, it is known that the association of two *Pf*LDH monomers to form a dimer is strong whereas the association of two dimers to form a tetramer is weak (Shoemark, 2000). Thus an alternative way an inter-subunit disulphide bond could form would be if the *Pf*LDH structure disassociated at the tetrameric interface (Figure 7.17) and the two dimers re-associate by forming a disulphide bond between two solvent exposed C183 residues.

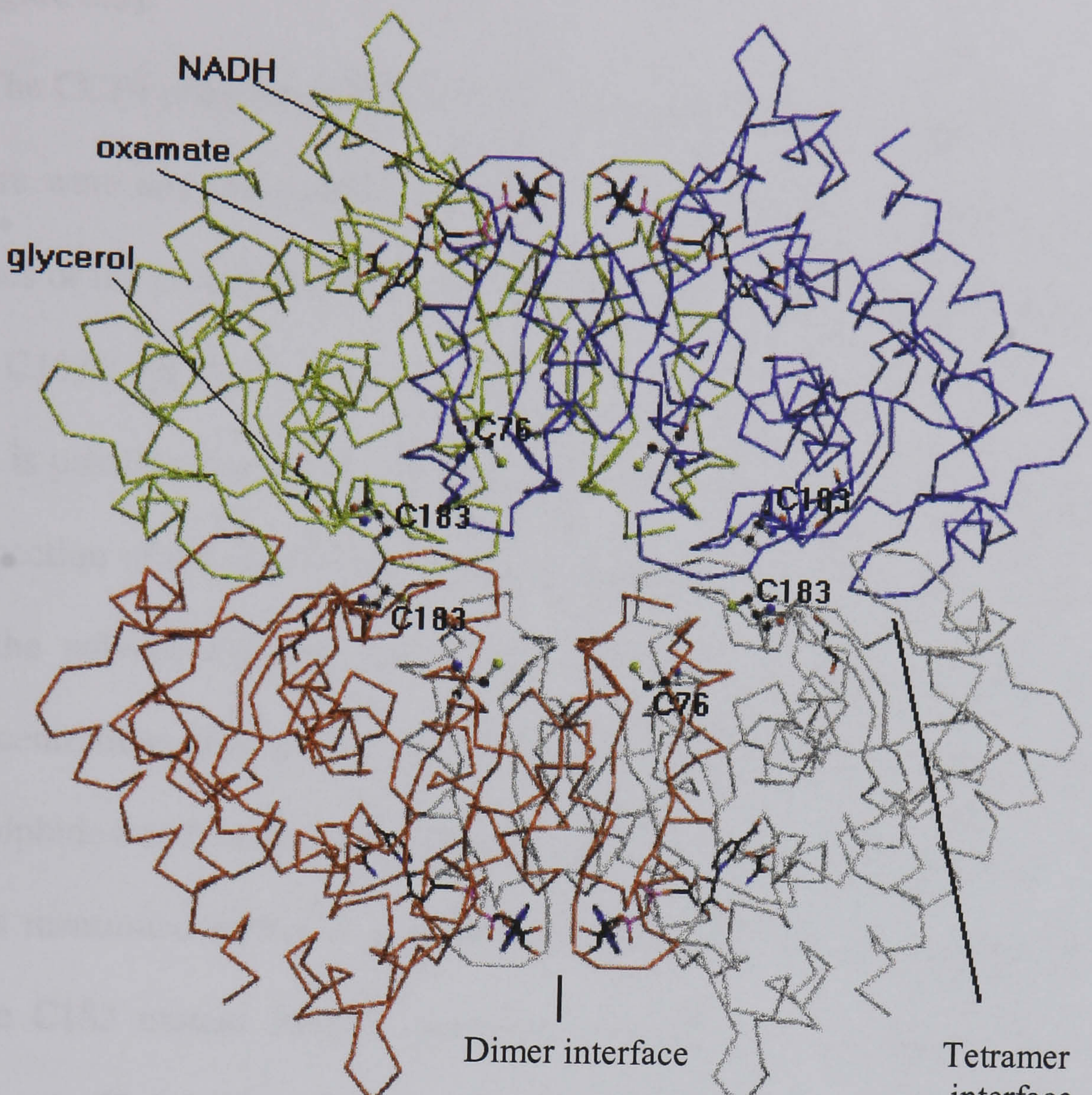


Figure 7.17 Tetrameric structure of His-tagged WT *Pf*LDH formed by the I222 symmetry in the crystal. Produced with *Molscript* (Kraulis, 1991).

Figure 7.18 Enlarged view of a section of the tetrameric structure of His-tagged WT *Pf*LDH. Image produced with *Quanta*.

Surprisingly, of the three C183 mutations only the valine substitution appeared to very slightly increase the stability of *Pf*LDH against oxidation (Figure 6.3).

The CCP4 program *CONTACT* was then used in order to ascertain whether there were any inter-subunit contacts/steric clashes in the various oligermic states of the protein that might help to explain the slight oxidative stability of the C183V *Pf*LDH mutant. However, no such contacts were found.

It is possible that C183 and C76 do not form a disulphide bond. On closer inspection of the *Pf*LDH structures a cysteine residue (131) was found to be at the solvent exposed surface of the subunit. At high enough protein concentrations, this residue may be involved in the formation of inter-subunit disulphide bonds, and could lead to the affects described in Chapter 6.

As mentioned earlier, a glycerol molecule was found to be present in all three C183 mutant *Pf*LDH structures (it is also present in the wild type enzyme). The ordered glycerol has either soaked into the crystal from the cryoprotectant solution or has come from the original protein storage buffer. The molecule binds at the subunit interfaces of the tetrameric structure (Figures 7.17 and 7.19).

The glycerol molecule's location helps to explain the stabilising influence of 10% glycerol on dimer subunit association reported by Shoemark (2000). Adams *et al.* (1973) found an exchangeable anion binding site at the tetramer subunit boundary of dogfish M_4 L-LDH that could bind citrate, sulphate, oxamate or pyruvate. Moreover, Hall *et al.* (1995) found that sucrose had a stabilising effect on the folding of rabbit muscle LDH and on its dimeric association. Therefore it is presumed that sucrose may stabilise rabbit muscle LDH in a similar manner that glycerol stabilises *Pf*LDH.

Figure 7.19 *SIGMAA* weighted $2|F_{obs}|-|F_{calc}|$ map (contoured at 1.5σ) from the final C183A *Pf*LDH model. Electron density corresponding to a glycerol molecule was seen in the same position in all three *Pf*LDH mutant models (the asterisks' denotes water molecules). Image produced using *O*.

7.8 Conclusions

Although the C183 mutations were far from the active site of the *Pf*LDH enzyme, the mutations still lead to a substantial decrease in activity. It is unclear as to how a mutation at the subunit interface would affect the active site architecture. However, it is probable that the decrease in enzymatic rate is the result of a structural change at the dimer subunit interface which leads an increase in the distances between the amino acid residues involved in catalysis and substrate binding and the substrate and cofactor molecules.

It difficult to determine whether the residues C183 and C76 form a disulphide bond between subunits. However, a solvent exposed cysteine residue was found. This residue, C131, could lead to the formation of inappropriate disulphide bonds and may help to explain the atmospheric

instability of the enzyme. Thus, mutagenesis of C131 may lead to a more stable form of the *Pf*LDH enzyme.

A glycerol molecule is present at the subunit interface of the C183 mutant (and wild type) *Pf*LDH structures. This may help to explain the stabilising influence of 10% glycerol on subunit association in *Pf*LDH. It may be possible to further enhance the stabilisation of the oligomeric state of the enzyme by using other small solutes such as sucrose and glucose.

Chapter 8

Crystallisation and Preliminary Characterisation of the Structure of a Recombinant Triple Mutant Pig M_4 L-LDH

8.1 Introduction

8.1.1 Mammalian LDH Isoenzymes

Vertebrate LDH's consist of four catalytically independent, tetrahedrally arranged subunits. Each subunit is approximately 36kD. Three forms of subunit are expressed in mammals, termed H, M and X. The X form appears to be restricted to postpubertal testes. The distribution of the H and M isoenzymes in aerobic and anaerobic forms reflects their different affinites for lactate and pyruvate (Holbrook *et al.*, 1975). The isoenzymes have evolved with common catalytic mechanisms but different susceptibilities to inhibition by NAD^+-pyruvate to suit their different metabolic roles. H_4 (consisting of four H subunits) predominates in the heart. The H_4 LDH catalyses the 'reverse' reaction (pyruvate to lactate) so that lactate from the tissues can fuel the heart. M_4 is the major LDH isoenzyme in anaerobic skeletal muscle.

The different subunits can combine to form mixed or hybrid tetramers (H_4, MH_3, M_2H_2, M_3H, and M_4), so it seems likely that the energy of inter-subunit contact is independent of isoenzyme type. An amino acid sequence comparison of the 12 known vertebrate reveals that the same isoenzymes from different species are more homologous than different isoenzymes from the same species. This suggests that the vertebrate isoenzymes are ancient and have remained distinct since the evolution of vertebrates.

8.1.2 A Recombinant Triple Mutant Pig M_4 LDH

Attempts to target LDH for specific drug design (e.g. *Pf*LDH for antimalarials) required a good model for the human LDH structures. The pig LDH isoenzymes were used as models, as the isoforms share considerable sequence identity. The pig and human LDH enzymes have a sequence identity of 97% and 92% between H and M isoforms respectively (Figure 1.4). Furthermore, in the pig M_4 LDH structure all 23 differing residues are all on the surface of the protein away from the active site (El Hawarani *et al.*, 1996). Therefore, the pig M_4 LDH serves as an excellent model for human M_4 LDH. (Very recently the structures of the human M and H forms of LDH have been solved (Read *et al.*, submitted for publication) and it was shown that the human and pig M subunits have an r.m.s.d of just 0.56Å.)

In order to carry out antimalarial selectivity trials a large amount of pig M_4 LDH was required. This was achieved by overproduction of the enzyme in *E.coli*: The pig lactate dehydrogenase A cDNA was ligated into pKK223-3 and the plasmid was subsequently transformed into *E.coli* TG2 cells (Otto Manhertz, unpublished results). However, during the synthesis of the recombinant three sets of non-silent mutations were introduced into the gene. This lead to the production of a recombinant pig muscle LDH with three amino acid mutations; S202N, V204I and F332L (Figure 8.5). The genetic mutations were eventually repaired, allowing the overproduction of recombinant 'wild type' LDH.

The enzymatic activties of the various forms of the pig muscle LDH were compared (Otto Manhertz, unpublished results) (Table 8.1):-

Protein	k_{cat} (s^{-1})	K_M pyruvate (mM)	K_i (mM)
Wild type (non-recombinant)	402	0.365	1.271
'Wild type' (recombinant)	369	0.423	0.571
Mutant	262	0.329	0.918

Table 8.1 Kinetic constants of the various forms of pig muscle LDH.

The enzymatic rate of the recombinant triple mutant pig M_4 LDH is about 60% that of the wild type non-recombinant protein. How could three mutations remote from the active site have such an effect on the kinetics of the enzyme? This question was addressed by obtaining the structure of the triple mutant LDH enzyme.

8.2 Protein Purification and Crystallisation

8.2.1 Protein Expression

The *E.coli* cells, carrying the pKK223-3 plasmid harbouring the pig A L-LDH gene, were selected by growing the cells in rich media containing 1μg/ml ampicillin.

A glycerol stock of the E.coli cells was streaked out onto 1.6% LB agar and grown overnight at 37°C. A single bacterial colony was used to innoculate 5ml L Broth. This was then grown at 37°C for 6 hours in a shaking incubator. The 5ml culture was then used to innoculate 500ml L Broth. When the culture had reached an A_{600nm} of 0.5 (mid-log phase of growth), protein expession was induced by the addition of 1mM IPTG. The cells were then allowed to grow for a further 2 hours.

8.2.2 Protein Purification

The cells were harvested by centrifugation at 4,000rpm for 20 minutes at 4°C. A suspension of the cells was prepared using a minimal volume of 50mM TEA pH 6.0 prior to sonication on ice, using 3 repeats of 15 second medium-power bursts. The subsequent cell debris was then pelleted by centrifugation at 4°C for 20 minutes at 18,000rpm. The supernatant was then dialysed against 50mM TEA pH 6.0 overnight.

The protein was then purified using a procedure similar to that described by Wilks *et al.* (1992). The dialysate was spun at 18,000rpm for 20 minutes at 4°C in order remove any sediment. After dialysis, NADH was added to the protein at a final concentration of 200μM.

The dialysate was then loaded onto an affinity column packed with 100ml oxamate-Sepharose (Sigma) which had been pre-equilibrated with dialysis buffer, containing 100μM NADH. After unbound protein had been eluted with column buffer, the triple mutant pig muscle L-LDH was eluted with 50mM TEA pH 9.0, 0.25M NaCl. LDH activity was determined by monitoring the drop in NADH absorbance at 340nm when the LDH sample was added to an assay mixture at 25°C (Figure 8.1). The assay mixture consisted of 0.2mM NADH, 1mM sodium pyruvate buffered with 50mM TEA pH 7.2.

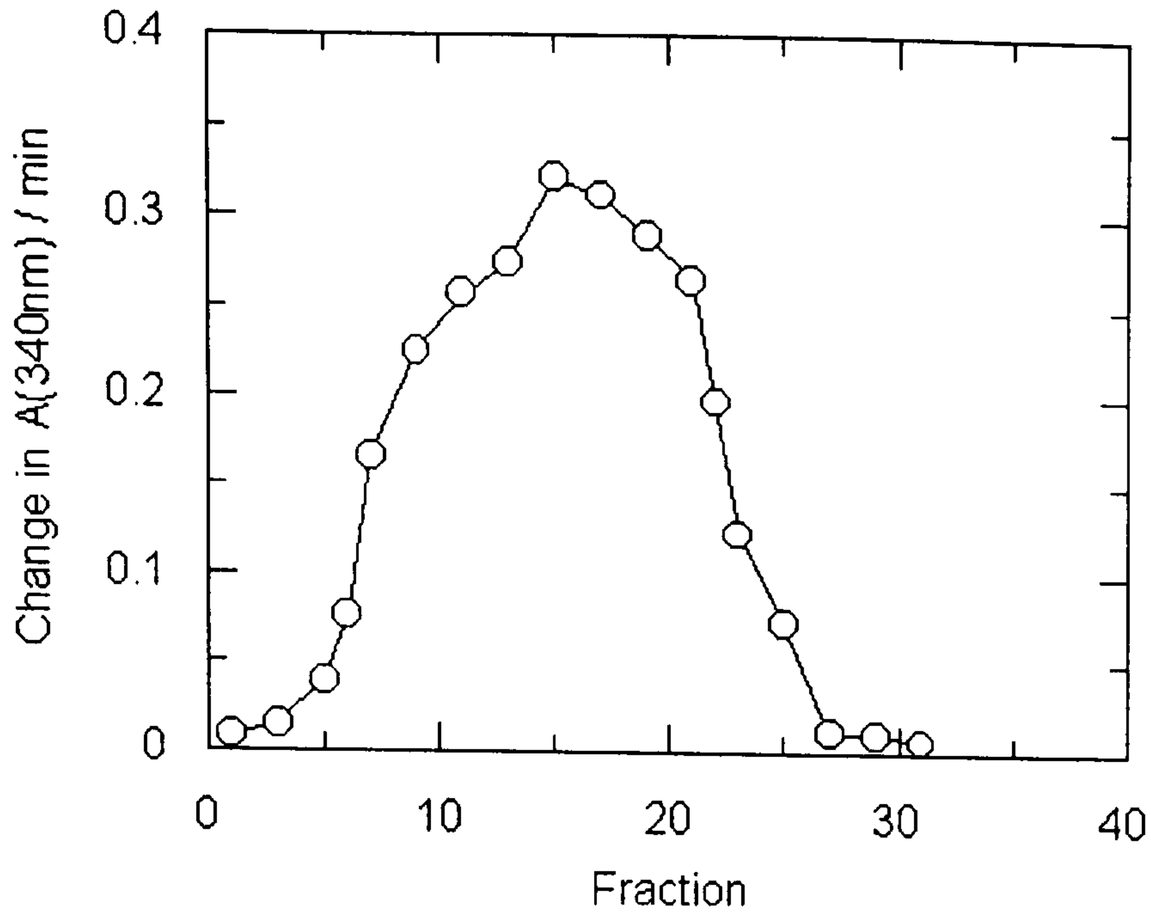


Figure 8.1 Plot of bound fraction eluting from the oxamate-Sepharose column using 0.25M NaCl.The elution profile was followed by measuring the drop in NADH absorbance at 340nm when the LDH sample was added to an assay mixture at 25°C.

The eluate was centrifuged at 18,000rpm for 20 minutes at 4°C in order remove any insoluble material. Then the supernatant was dialysed against 50mM TEA pH 7.5. The protein was then loaded onto a 150ml Q-Sepharose column which had been pre-equilibrated with 2 column volumes of 50mM TEA pH 7.5. The protein was then eluted with a 500ml 0-0.5M NaCl gradient. The LDH protein eluted at around 0.2M NaCl. Again LDH activity was by measured by monitoring the drop in NADH absorbance at 340nm on addition of the protein sample (Figure 8.2).

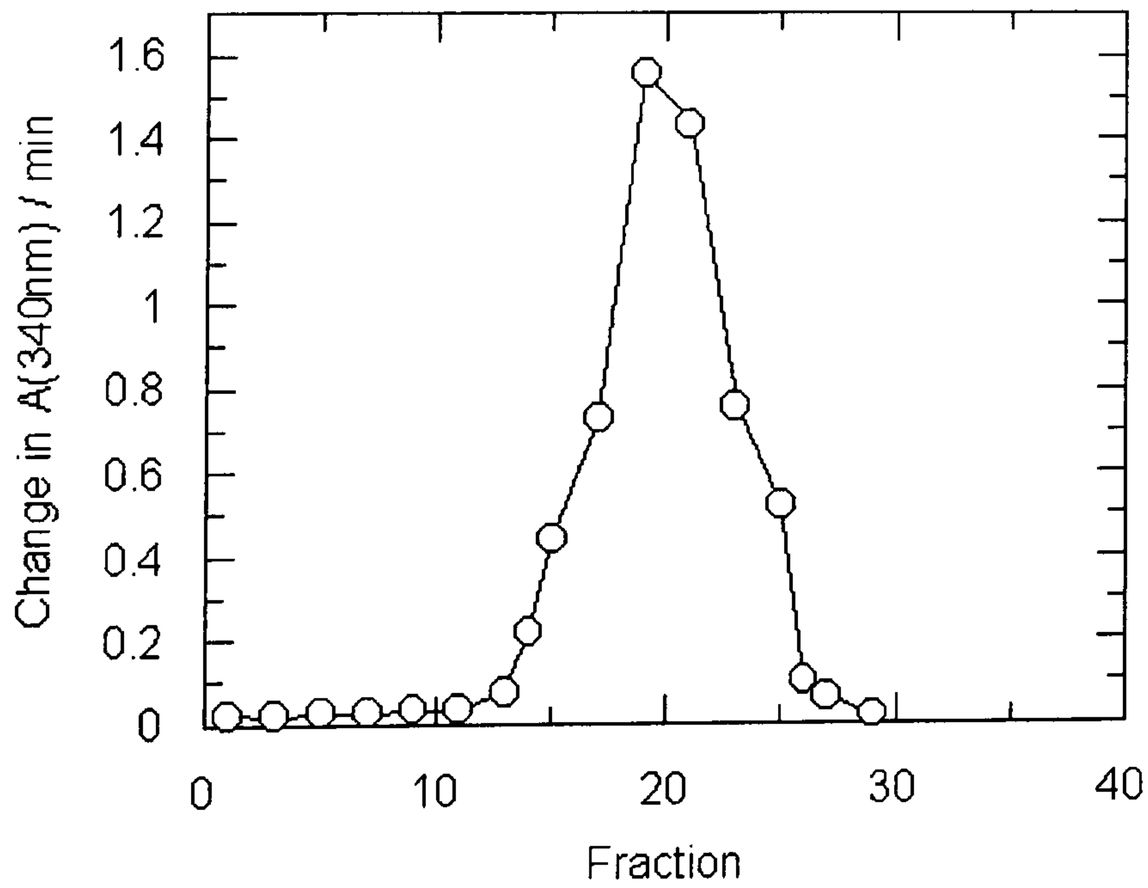


Figure 8.2 Plot of bound fraction eluting from the Q-Sepharose column using a 0-0.5M NaCl gradient. The LDH protein eluted at around 0.25M NaCl. The elution profile was followed by measuring the drop in NADH absorbance at 340nm when the LDH sample was added to an assay mixture at 25°C.

SDS-PAGE analysis showed that the protein had been purified to homogeneity (gel not shown). The protein was then filter sterilised using a 0.22μm syringe driven filter, and then stored at 4°C in 40% (w/v) ammonium sulphate.

8.2.3 Crystallisation

After purification, the protein concentration of the sample was measured by recording the absorbance spectrum of the sample. The extinction coefficient at 280nm, of a 1% solution of pig muscle L-LDH was taken to be 1.39 (Jecsai, 1962). If the A_{280nm}/A_{260nm} measured above 1.9, the purified protein sample was dialysed against 0.5g/l activated charcoal in order to remove any bound cofactor. The samples were then concentrated and buffer exchanged against milli Q water using 30kD MWCO Centricon filters.

Crystals of the pig M_4 L-LDH/NADH/oxamate ternary complex were obtained by using a modified method to that proposed by Stinson and Gutfreund (described in Hackert *et al.*, 1973). Diamond-shaped prism like ternary crystals grew at 4°C using the hanging drop vapour diffusion method against a well containing 1.73-1.78M $(NH_4)_2SO_4$. The mother liqour contained 3.5-4.5 mg/ml protein, 20mM TEA pH 7.9-8.1, 240μM NADH and 24mM sodium oxamate.

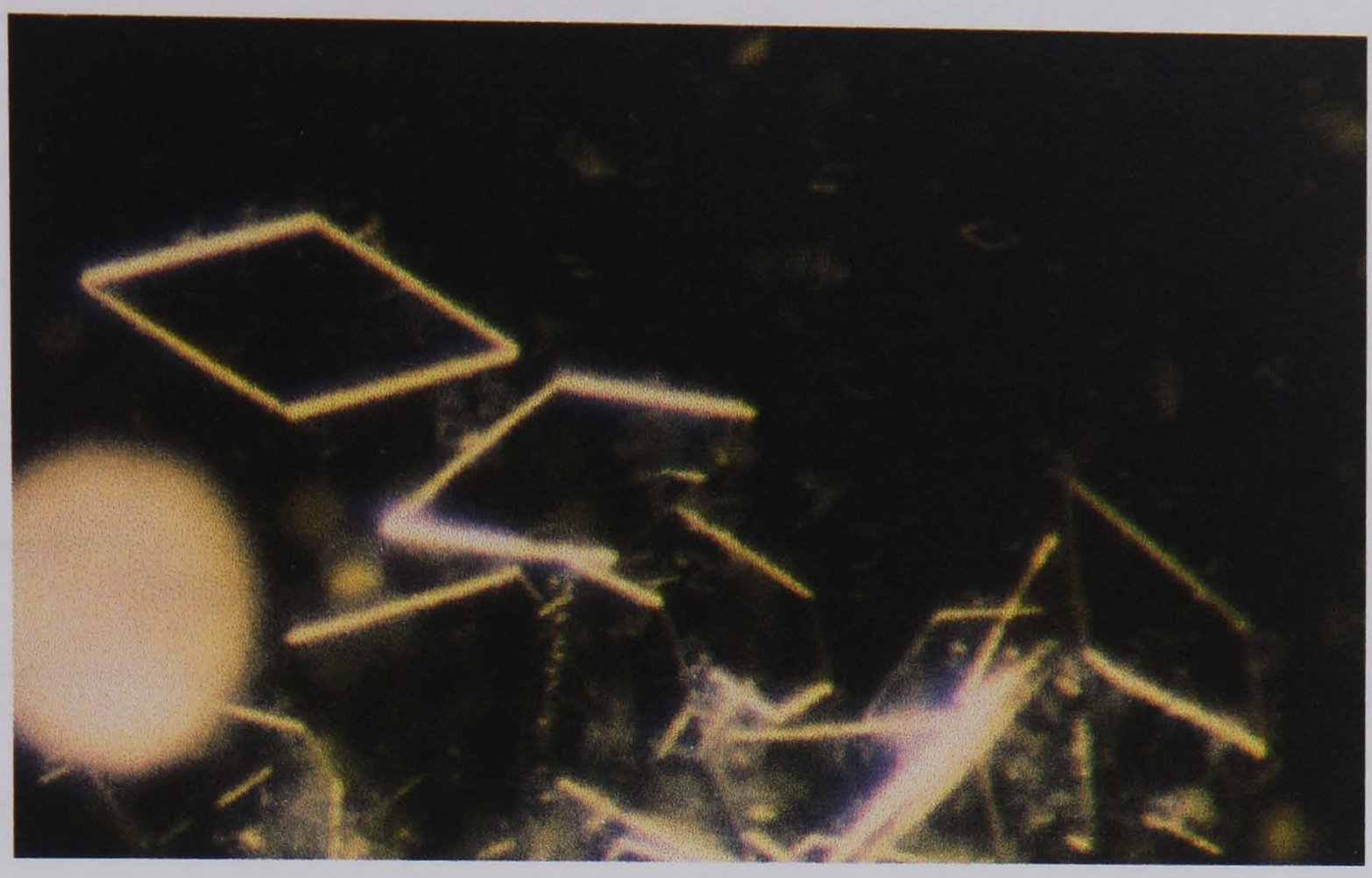


Figure 8.3 Ternary complex crystals of the S202N/V204I/F332L pig muscle L-LDH.

8.3 Data Collection

The in-house rotating anode X-ray source (λ=1.5418Å) was used to collect diffraction data at 100K from a single pig muscle L-LDH crystal. The data statistics are summarised in Table 8.2.

Cryoprotectant solution	1.8M $(NH_4)_2SO_4$, 20mM TEA pH 8.0, 240μM NADH, 24mM oxamate, 30% glycerol.
Unit cell	a=59.13, b=135.35, c=84.79, $\alpha=\beta=\gamma=90°$
Space group	$P2_12_12$
Total number of unique reflections	14,695
Mean Redundancy	2.64
Resolution	2.9Å
Mosaicity	1.62°
Completeness - Overall (20.00-2.9Å) Outer shell (3.12-2.9Å)	92.9% 85.3%
I / σ (I) - Overall Outer shell	6.66 2.60
R_{merge} - Overall Outer shell	12.6% 28.9%

Table 8.2 Diffraction data collection statistics of the pig M_4 LDH mutant triple mutant ternary complex crystal.

$V_M = 4.7$ Å^3/Da $\Rightarrow$ 2 molecule/a.s.u.

The data were found not to be complete. However, the completeness of the diffraction data could not be improved upon as other pig M_4 LDH triple mutant crystals did not diffract to a high enough resolution (~3.5Å).

The pig muscle L-LDH triple mutant crystallises in the $P2_12_12$ space group, with two molecules in the asymmetric unit. This implies that one of the tetramer 2-fold rotation axes is parallel to the crystallographic 2-fold axis (Hahn, 1995; Hackert *et al.*, 1973).

The diffraction data were then processed and reduced.

8.4 Solving the Pig M_4 LDH Triple Mutant Structure

The structure was solved by molecular replacement using the *AMoRe* program. The monomeric structure of a ternary complex of wild type pig M_4 LDH (Dunn *et al.*, 1991) (with the NADH and oxamate removed) was used as a search model.

The top 10 rotation peaks obtained (Table 8.3) were used in the translation function (Table 8.4), to identify the first translation solution. This was then fixed and used to find the translation solution for the second molecule in the asymmetric unit (Table 8.5). The two solutions obtained clearly stand out from the background. Rigid body refinement was preformed via 'fiting' in *AMoRe*, to maximise the correlation between the calculated and observed structure factor amplitudes, giving the final positions of the solutions (Table 8.6).

Radius of integration: 25Å. Cell model: 125Å × 125Å × 125Å.

Resolution cutoff: 10.0-4.0Å.

Solution Number	**Rotation of Eulerian angles (°)**			**Correlation coefficient (%)**
	α	β	γ	
Solution 1	45.57	74.94	344.48	17.9
Solution 2	72.70	73.97	343.60	16.2
Solution 3	70.11	52.81	29.65	10.9
Solution 4	73.50	48.97	20.61	10.8
Solution 5	137.45	90.00	190.52	9.3
Solution 6	42.55	90.00	9.63	9.3
Solution 7	55.51	89.00	313.75	8.9
Solution 8	123.50	90.00	134.68	8.9
Solution 9	177.06	52.16	147.40	8.9
Solution 10	172.33	82.43	179.23	8.6

Table 8.3 The rotation solutions from *AMoRe* for the pig M_4 LDH mutant triple mutant ternary complex using the monomer of the wild type pig M_4 LDH as the search model.

Solution Number	**Rotation of Eulerian angles (°)**			**Translation (fractions of the unit cell)**			**Correlation coefficient**	**R-factor**
	α	β	γ	**x**	**y**	**z**	(%)	(%)
Solution 1	45.57	74.94	344.48	0.1092	0.3749	0.2331	30.6	48.0
Solution 2	72.70	73.97	343.60	0.2269	0.4081	0.0688	30.3	47.3
Solution 3	70.11	52.81	29.65	0.3221	0.0337	0.4701	9.1	54.2
Solution 4	73.50	48.97	20.61	0.4577	0.3482	0.0220	8.6	53.9
Solution 5	137.45	90.00	190.52	0.3817	0.4289	0.0415	9.9	53.6
Solution 6	42.55	90.00	9.63	0.1213	0.4298	0.4614	9.6	53.8
Solution 7	55.51	89.00	313.75	0.0988	0.0336	0.0574	9.8	53.6
Solution 8	123.50	90.00	134.68	0.9909	0.0789	0.1070	10.1	53.6
Solution 9	177.06	52.16	147.40	0.3545	0.3152	0.4310	7.5	53.9
Solution 10	172.33	82.43	179.23	0.3346	0.2731	0.2036	8.8	54.0

Table 8.4 The initial translation solutions from *AMoRe* for the pig M_4 LDH mutant triple mutant ternary complex using the monomer of the wild type pig M_4 LDH as the search model.

Solution Number	Rotation of Eulerian angles (°)			Translation (fractions of the unit cell)			Correlation coefficient	R-factor
	α	β	γ	x	y	z	(%)	(%)
Fixed	45.57	74.94	344.48	0.1092	0.3749	0.2331	30.6	48.0
Solution 1								
Solution 2	72.70	73.97	343.60	0.7238	0.9082	0.0685	61.0	36.8
Solution 3	70.11	52.81	29.65	0.8279	0.5088	0.9054	21.7	50.8
Solution 4	73.50	48.97	20.61	0.7571	0.0974	0.6986	21.6	50.9
Solution 5	137.45	90.00	190.52	0.2325	0.4284	0.3603	22.3	50.5
Solution 6	42.55	90.00	9.63	0.2667	0.9287	0.6403	22.2	50.4
Solution 7	55.51	89.00	313.75	0.5868	0.0179	0.8939	22.1	50.5
Solution 8	123.50	90.00	134.68	0.4157	0.8737	0.6066	22.2	50.4
Solution 9	177.06	52.16	147.40	0.0604	0.0514	0.7008	21.4	50.6
Solution 10	172.33	82.43	179.23	0.4442	0.9820	0.2375	21.6	50.3

Table 8.5 A search for the second translation solution for the pig M_4 LDH mutant triple mutant ternary complex from *AMoRe* after fixing the first solution and using the monomer of the wild type pig M_4 LDH as the search model.

Solution Number	Rotation of Eulerian angles (°)			Translation (fractions of the unit cell)			Correlation coefficient	R-factor
	α	β	γ	x	y	z	(%)	(%)
Solution 1	45.78	75.22	344.94	0.1100	0.3741	0.2314	68.8	32.9
Solution 2	72.51	74.88	344.68	0.7251	0.9091	0.0679	68.8	32.9

Table 8.6 The final solutions after 'fiting' from *AMoRe* for the pig M_4 LDH mutant triple mutant ternary complex using the monomer of the wild type pig M_4 LDH as the search model.

The coordinates of the wild-type pig M_4 LDH search model were then rotated and translated to the position in the unit cell determined by molecular replacement.

The dimer in the asymmetric unit was then constructed in the graphics program *O* using the monomer and symmetry equivalents. As with the wild type enzyme, the $P2_12_12$ packing of the triple mutant pig muscle LDH produces a tetramer (Figures 8.4 and 8.5).

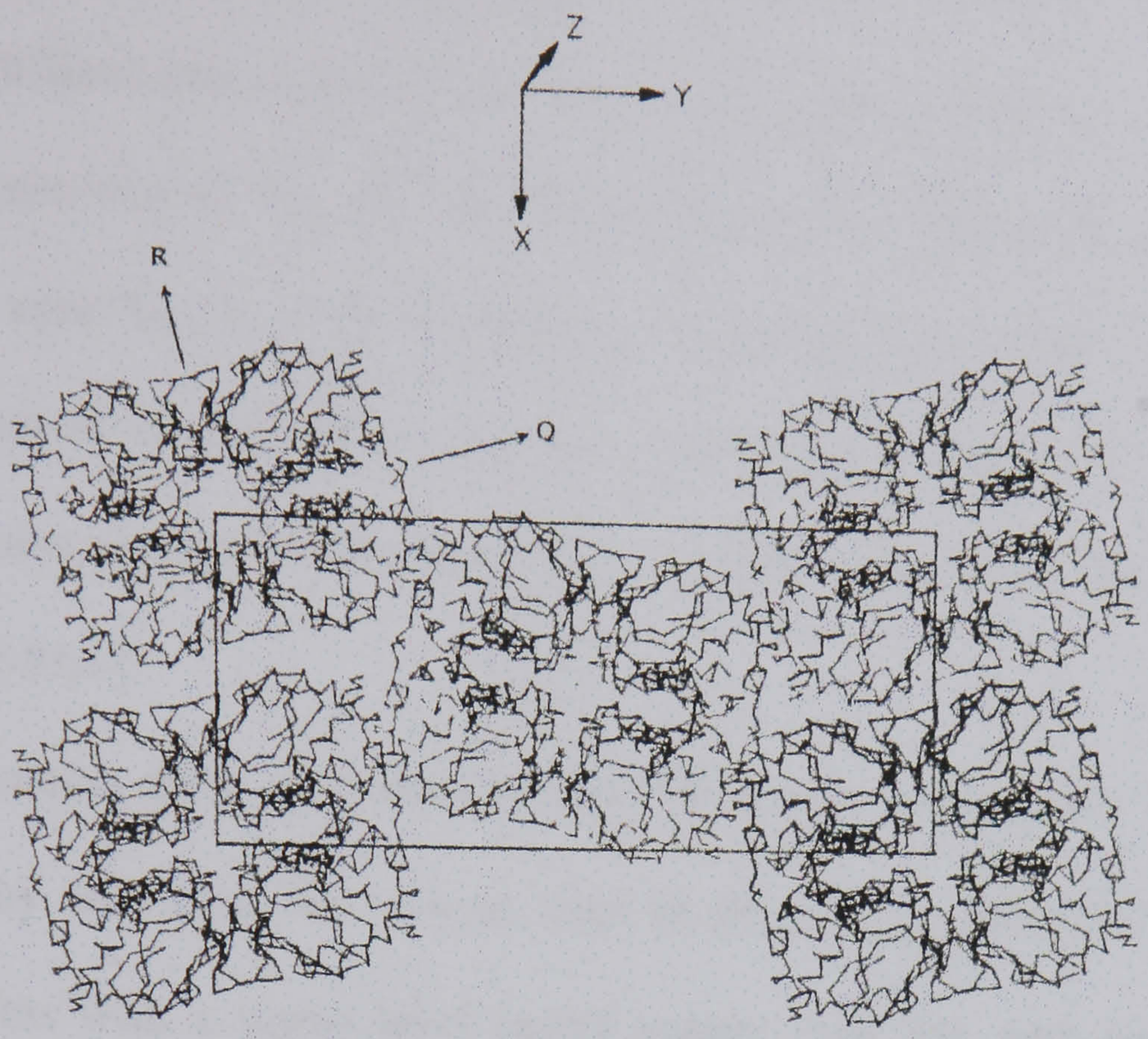


Figure 8.4 The packing arrangement of the tetramer in the pig M_4 LDH crystals. Viewed down the the Z axis, with the unit cell outlined. (The P, Q, R are described by Rossmann *et al.*, 1971).

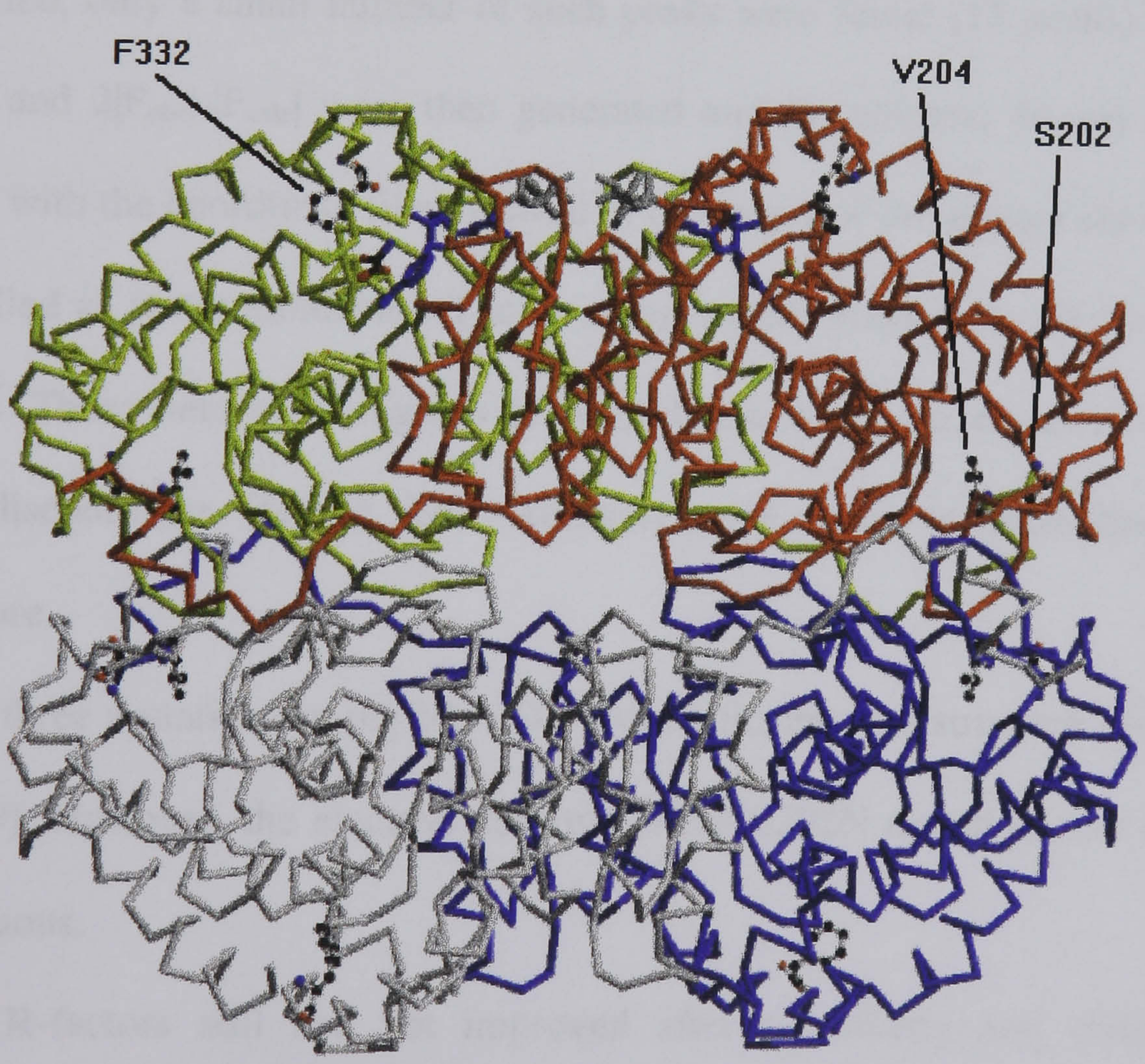


Figure 8.5 Tetrameric structure of the wild type Pig M_4 LDH formed by the $P2_12_12$ symmetry in the crystal. The residues shown were accidently mutated. Produced with *Molscript*.

8.5 Crystallographic Refinement and Model Building

The low resolution and incompleteness of the data made refinement of the structure difficult owing to the low observation to parameter ratio.

After generating an R_{free} test set, the structure was subjected to a round of simulated annealing in CNS at a starting temperature of 2500K (with the B-factors reset to 30.0Å^2). The R_{cryst} value decreased substantially. However, the R_{free} value remained high (Table 8.7), this may suggest that the data were being over-fitted.

The structure then underwent overall B factor refinement. This was followed by searching for solvent sites in $|F_{obs}|-|F_{calc}|$ difference maps at positive sites with a sigma level cutoff greater than 3.0, with temperature factors less than 50Å^2 and within hydrogen bonding distance (3.5Å). As expected, only a small number of such peaks were found (15 peaks). $|F_{obs}|-|F_{calc}|$ and $2|F_{obs}|-|F_{calc}|$ were then generated and the electron density maps, along with the coordinate files, viewed in *O*. Only 4 of the solvent sites were modelled as water molecules. The electron density maps were very poor in places: There were many regions of the molecule where the electron density was discontinous (Figure 8.6). Attempts were made at rebuilding the structure.

The three mutations in pig M_4 LDH were built into the structure (Figures 8.7-8.9). However, the electron density for the S202N mutation was rather ambiguous.

The R-factors still had not improved after re-building and conjugate minimisation of the model.

As there are two LDH subunits in the asymmetric unit, NCS (Non-Crystallographic Symmetry) constraints were applied during refinement. The rotation and translation matrix relating one molecule to the other was calculated using *lsqkab*. Maximum likelihood refinement was then carried out using *Refmac* version 5 (this has a bulk solvent correction alogarithm). Unfortunately, the R-factors and electron density maps did not improve.

Without over-fitting the data it was not possible to refine the structure further.

p No.	Refinement Procedure	Final R_{cryst} (%)	Final R_{free} (%)	Geometry bond lengths/ bond angles
1	*AMoRe*- rigid body fiting	32.9	-	-
2	Simulated annealing- 2500K	24.8	34.5	
3	Overall B-factor refinement	23.6	33.21	0.009Å/ 1.54°
4	*water_pick*	24.1	33.3	
5	Rebuild followed by minimisation	23.7	33.2	-
6	Overall B-factor refinement	23.6	33.3	0.010Å/ 1.56°
7	NCS averaging followed by *Refmac* 5	23.6	33.4	

Table 7.13 Summary of crystallographic refinement and rebuilding of the pig M_4 triple mutant LDH.

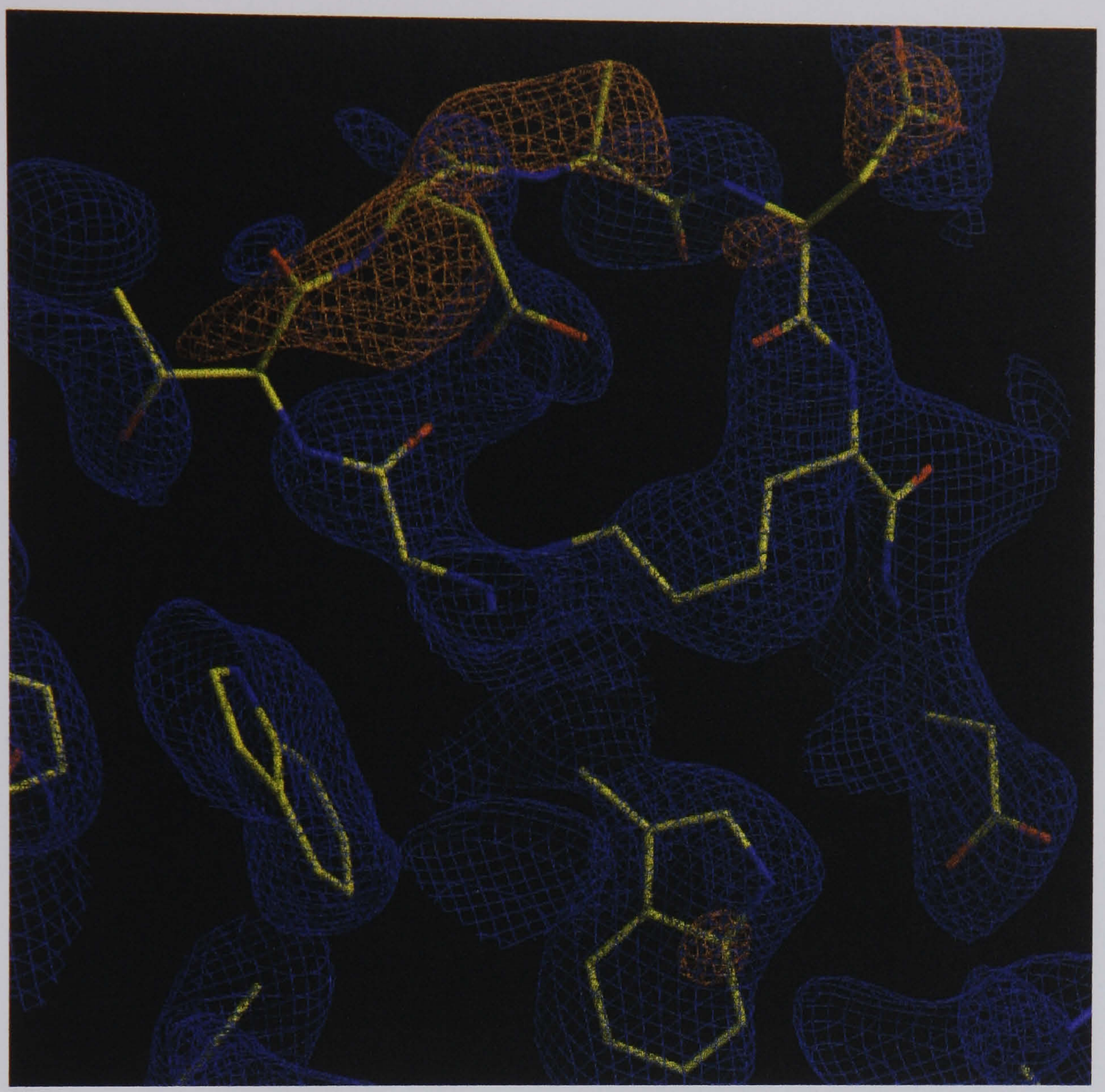
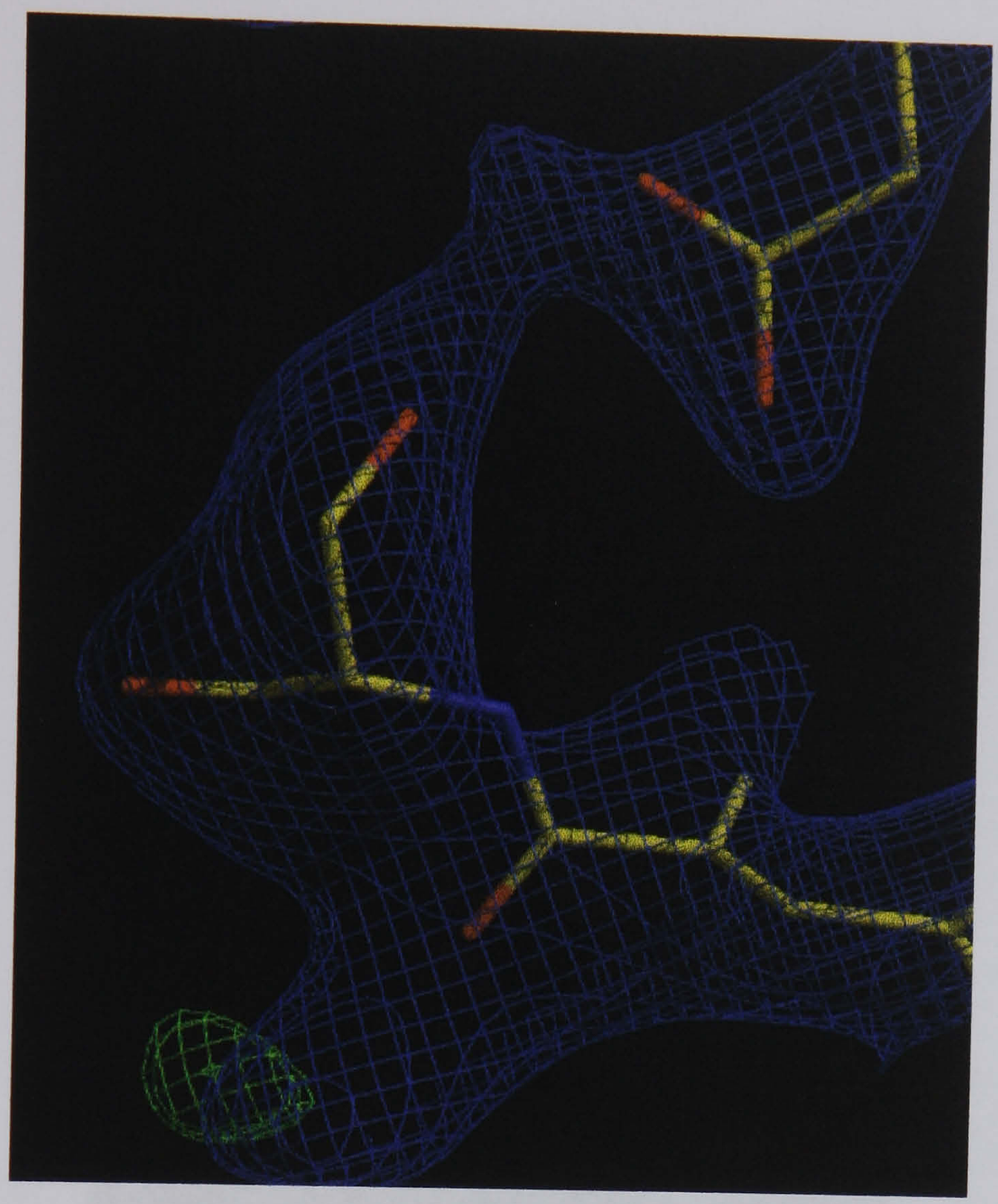


Figure 8.6 *SIGMAA* weighted $2|F_{obs}|-|F_{calc}|$ (purple, contoured at 2σ), and $|F_{obs}|-|F_{calc}|$ negative difference (brown, contoured at 3σ) maps. Some regions of the map have large breaks in electron density. Images produced using the *Xfit* program.

(a)



(b)

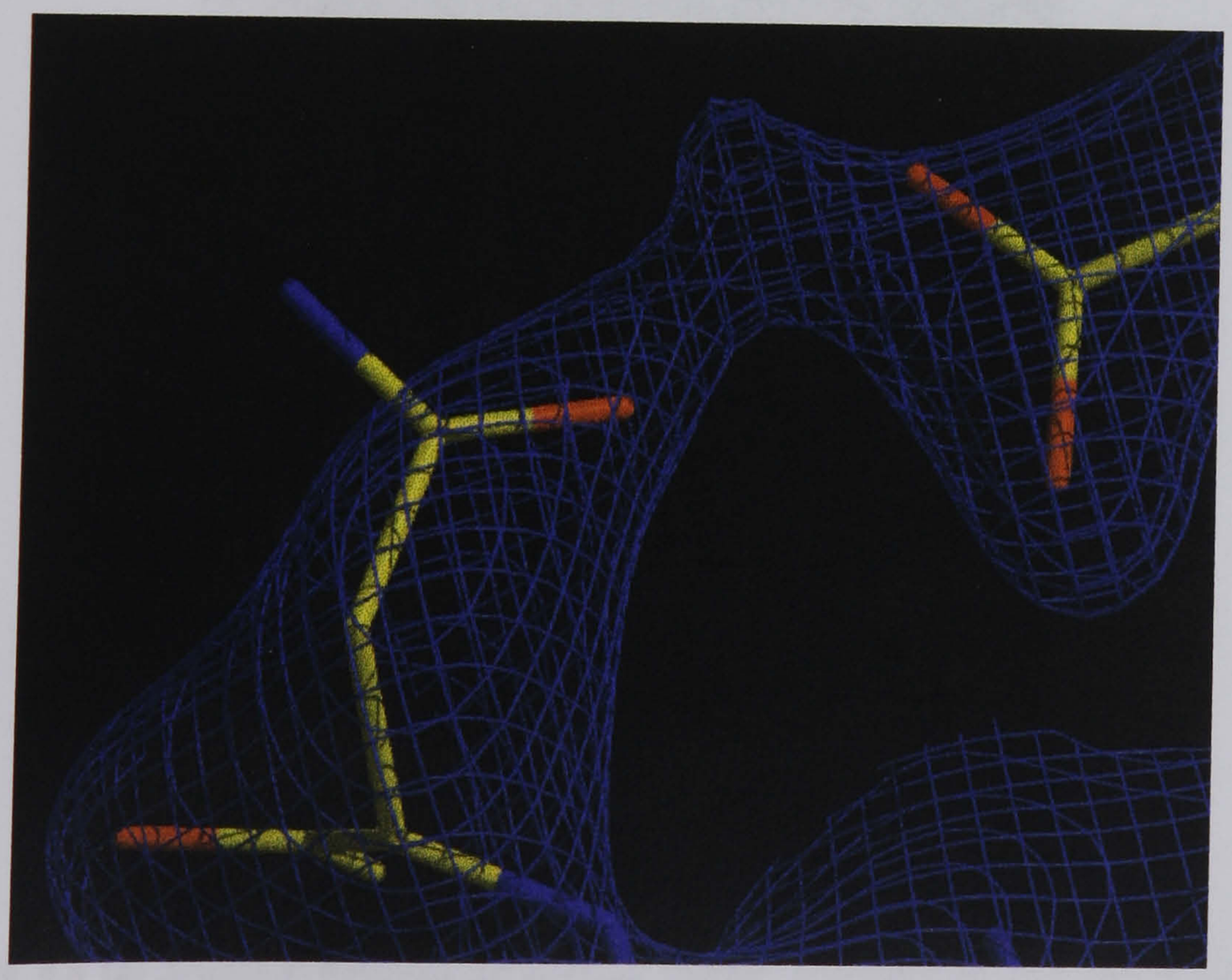
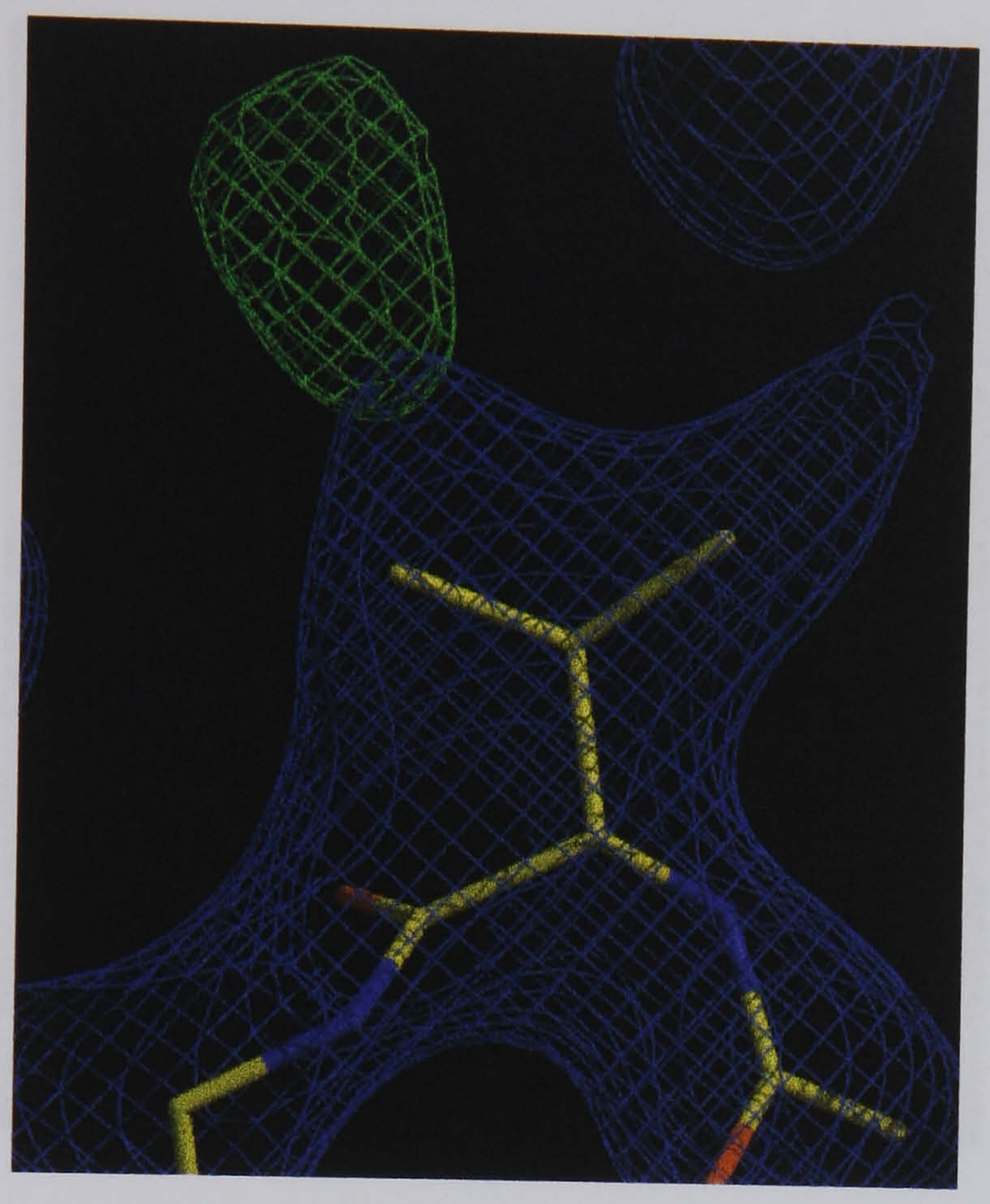


Figure 8.7 (a) *SIGMAA* weighted $2|F_{obs}|-|F_{calc}|$ (purple, contoured at 1σ), and $|F_{obs}|-|F_{calc}|$ positive difference (green, contoured at 3σ) maps. The S202N mutation is not clearly evident. (b) Substitution of the serine residue by asparagine. Images produced using the *Xfit* program.

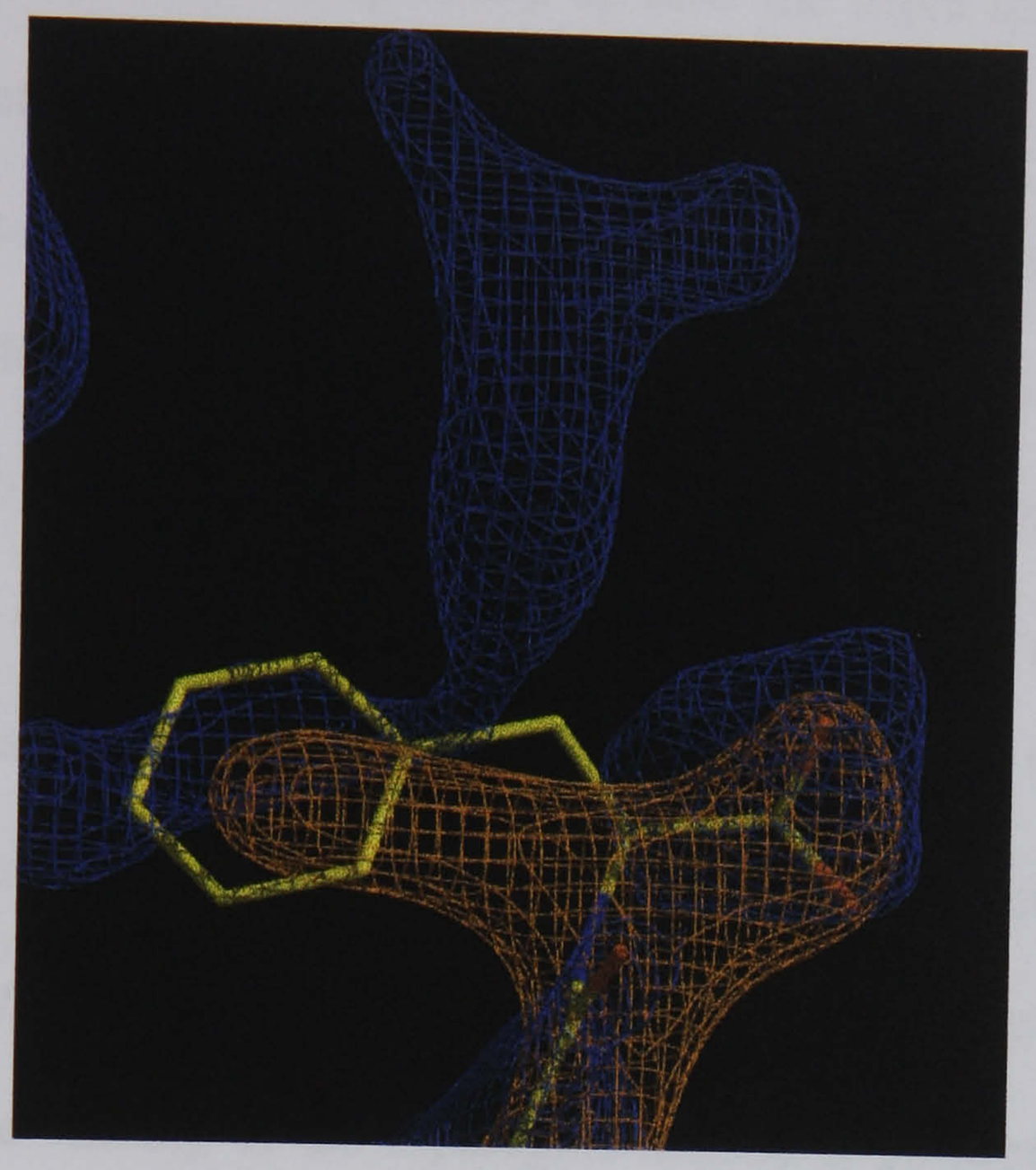
(a)



(b)

Figure 8.8 (a) *SIGMAA* weighted $2|F_{obs}|-|F_{calc}|$ (purple, contoured at 1σ), and $|F_{obs}|-|F_{calc}|$ positive difference (green, contoured at 3σ) maps. There is some positive density, suggesting that there is a branched residue at this locus. (b) Substitution of the valine 204 residue by isoleucine. Images produced using the *Xfit* program.

(a)



(b)

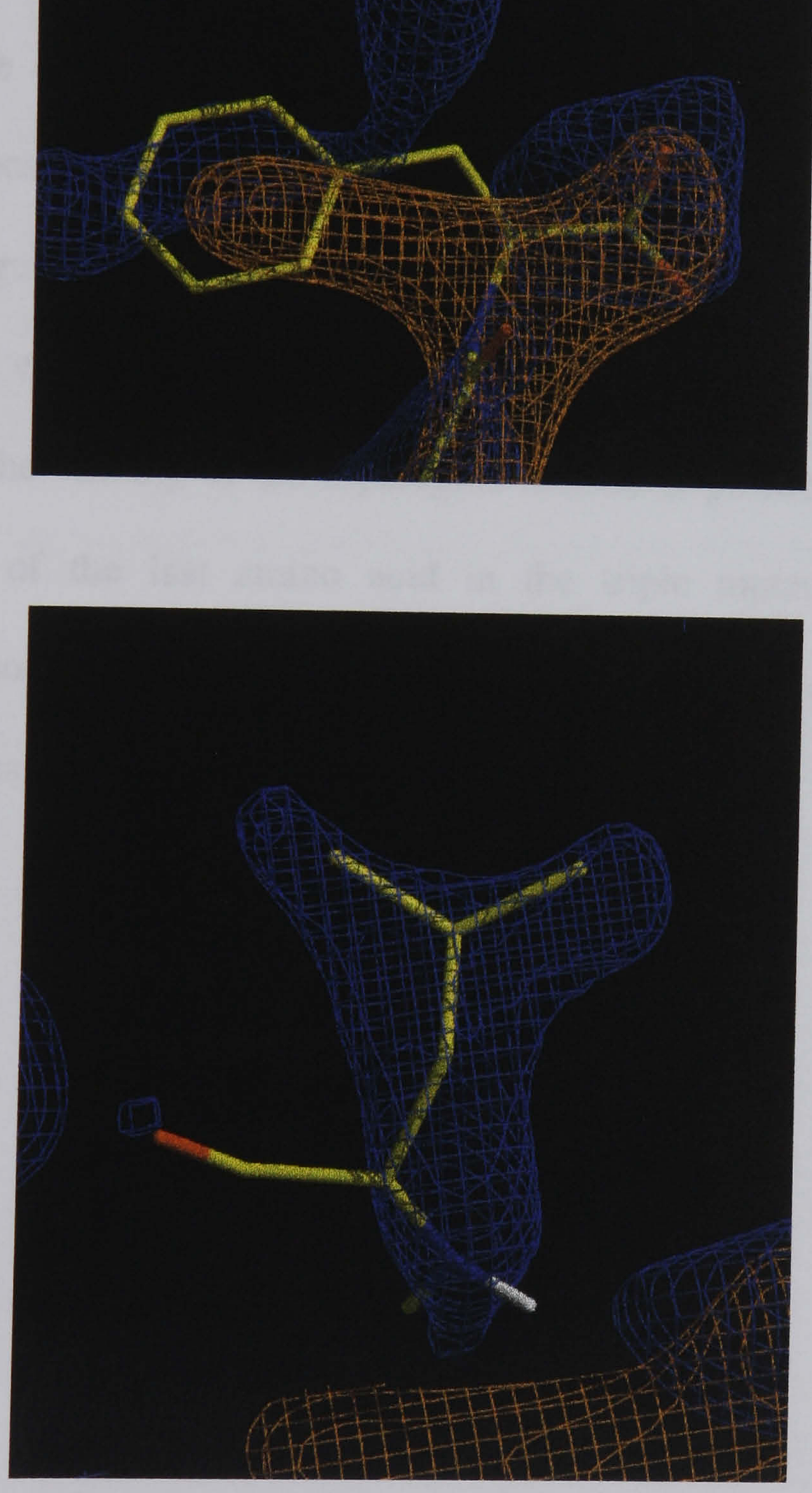


Figure 8.9 (a) *SIGMAA* weighted $2|F_{obs}|-|F_{calc}|$ (purple, contoured at 1σ), $|F_{obs}|-|F_{calc}|$ positive difference (green, contoured at 3σ), and $F_{obs}|-|F_{calc}|$ negative difference (brown, contoured at 3σ) maps. The negative difference map suggests that phenylalanine shoud not be present at this position. Moreover, the $2|F_{obs}|-|F_{calc}|$ maps shows that there is a branched residue at this locus. However, the electron density is discontinuous. (b) Substitution of the phenylalanine residue at position 332 by leucine. Images produced using the *Xfit* program.

8.6 Discussion

Although the quality of the structure of the triple mutant pig M_4 LDH is rather poor, it is possible to speculate on how the triple mutation in LDH could have an effect on enzyme activity.

The loss in activity of the mutant could be attributed to the destabilising change at residue 202. The serine at position 202 in the wild type pig M_4 LDH forms two hydrogen bonds with glutamate 309 (Figure 8.10). However, the asparagine residue at position 202 in the triple mutant structure could occur in a position where it only forms one hydrogen bond with the glutamate (Figure 8.11). This destabilisation is likely to be compounded by the change of valine at position 204 to the larger isoleucine residue as the residue is in the vinicity of the asparagine residue at position 202. Finally, the mutation of the last amino acid in the triple mutant, F332L, has presumbably no effect on the activity of the enzyme as the phenylalanine and leucine residues are remote from the other amino acids.

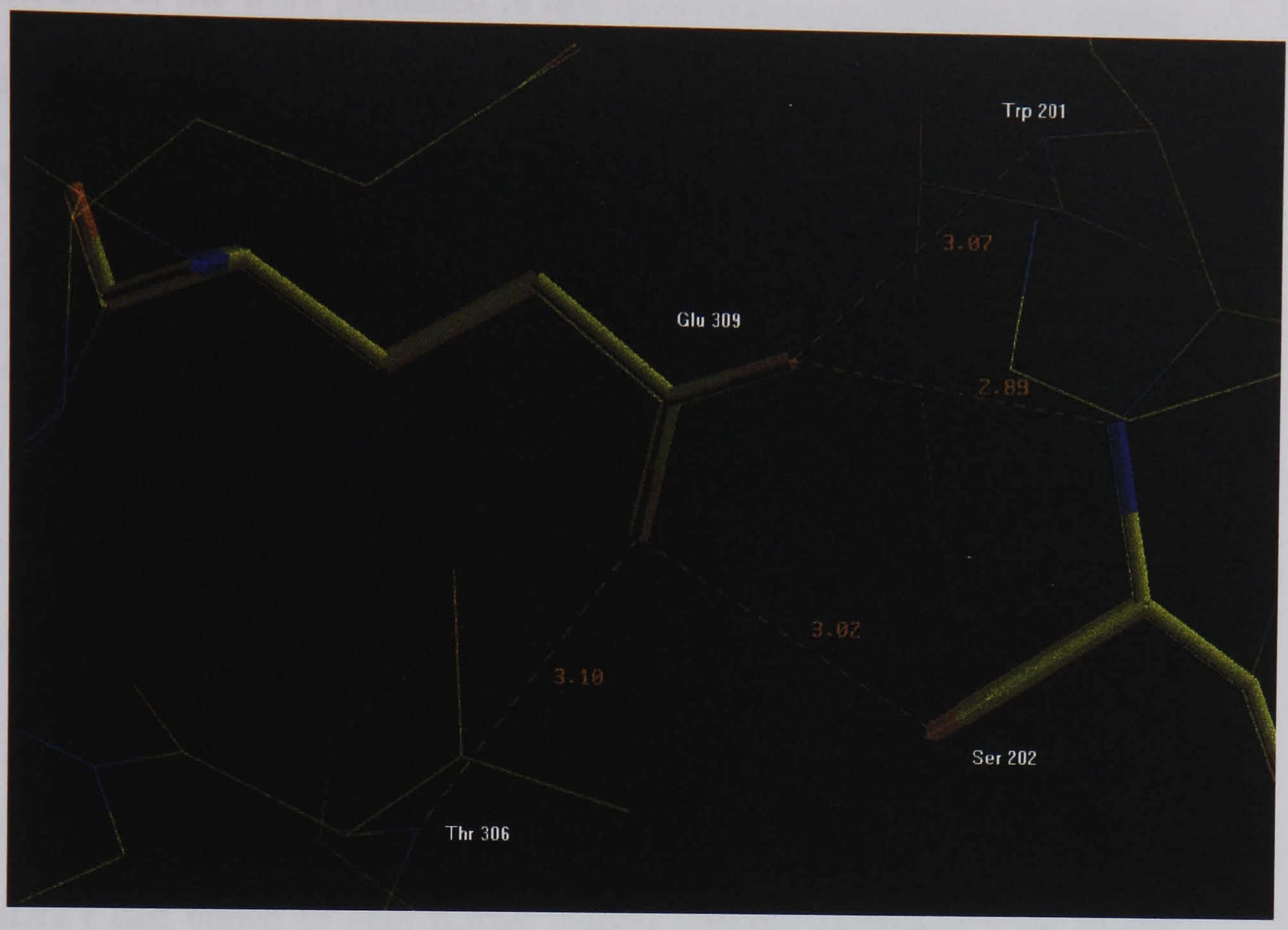


Figure 8.10 In the wild type pig M_4 LDH structure Ser 202 forms two hydrogen bonds with Glu 309. Produced in *O*.

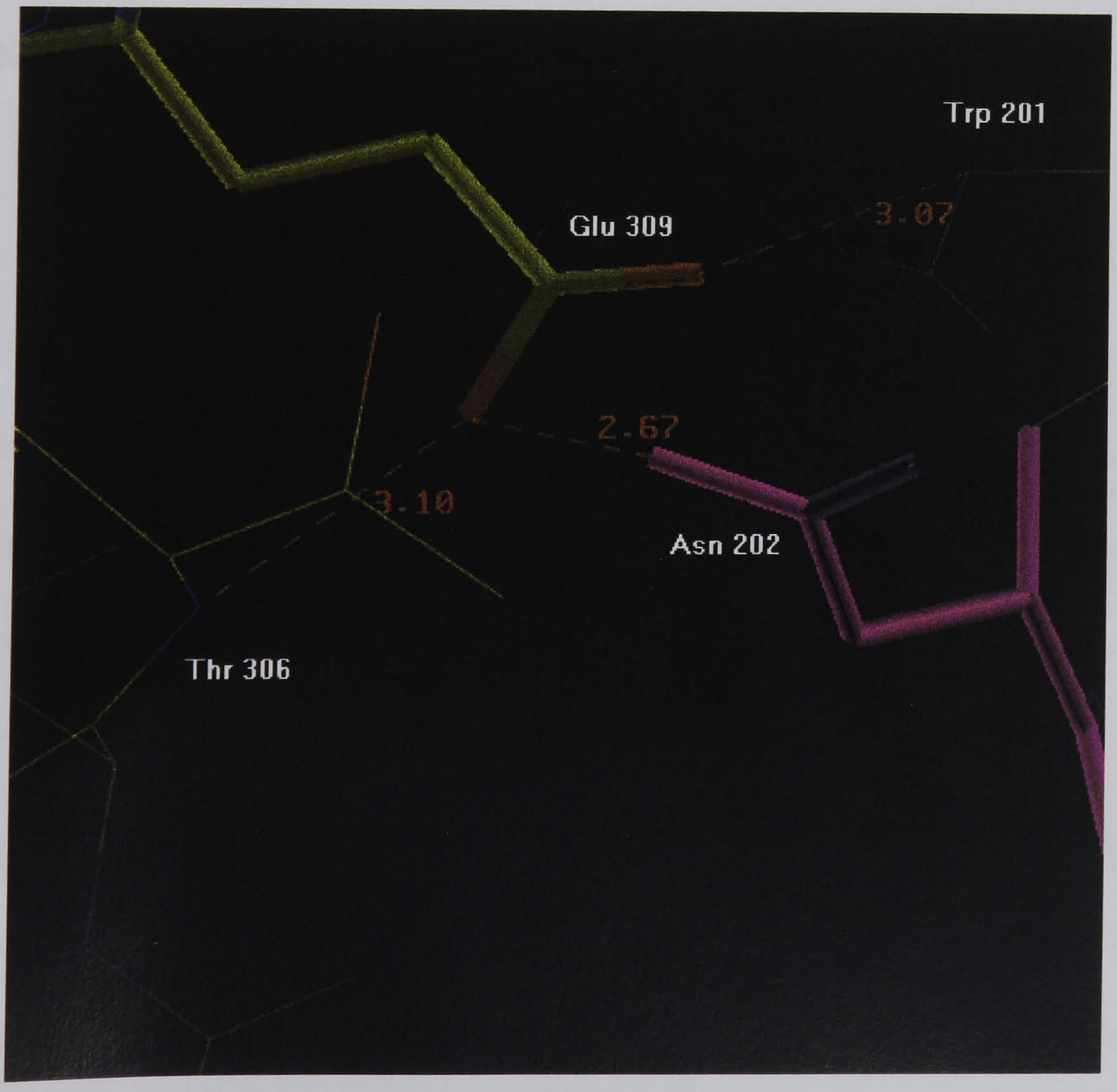


Figure 8.11 In the triple mutant pig M_4 LDH structure Asn 202 forms just one hydrogen bond with Glu 309. (Image taken from the current model of the triple mutant pig M_4 LDH) Produced in *O*.

Summary

The aim of the work presented in this thesis was to understand the kinetic and mechanistic properties of NAD^+-dependent D- and L- 2-hydroxyacid dehydrogenases and their mutants from a structural point of view.

L.bulgaricus D-2-hydroxyisocaproate dehydrogenase (D-HicDH) catalyses the NAD^+-dependent stereospecific and reversible reduction of a broad spectrum of aliphatic and aromatic 2-ketoacids. Previous site-directed mutagenic studies had shown that substitution of histidine at position 205 with glutamine leads to a dramatic increase in the catalytic efficiency of the enzyme. By elucidating the structure of the the wild type and mutant D-HicDH I had hoped to gain further insight into the molecular basis of the enzyme's mechanism, substrate specificity and stereospecificity.

Numerous crystal forms of the wild type and H205Q mutant D-HicDH were found. However, attempts at phasing the subsequent diffraction data using several molecular replacement programs were unsuccessful. The inability to solve the D-HicDH structure using some of the data sets may have been due to incompleteness of the data. However, the inability to solve the structure may suggest that there are substantial structural differences between the target and search molecules.

The multiple isomorphous replacement method was then used in an attempt to phase the D-HicDH data sets. Unfortunately, numerous attempts at preparing heavy atom derivatives had been unsuccessful. This was due to the poor diffraction by the soaked crystals, high mosaicity, nonisomorphism and / or low occupancy of the heavy atoms within the crystals.

The multiwavelength anomalous dispersion phasing method was then employed to try to solve the D-HicDH structure. In order to obtain a usable anomalous signal, methionine residues in the protein had to be substituted for selenomethionine. 100% Selenomethionyl incorporation into the protein was eventually achieved by transforming the plasmid harbouring the D-HicDH gene into a methionine auxotrophic cell strain. However, the subsequent Se-Met D-HicDH were not of sufficient quality for MAD phasing experiments.

The L-lactate dehydrogenase from *Plasmodium falciparum* is a target for antimalarial drug design. Previous kinetic experiments had shown the enzyme to be catalytically unstable against oxidation. This was thought to be the result of inappropriate inter-subunit disulphide bond formation. In order to stabilise the enzyme, one of the cysteine residues at the subunit interface was mutated to valine, alanine and serine. The mutations did not increase the oxidative stability of *Pf*LDH and in fact reduced the catalytic activity of the enzymes by varying degrees. The structures of these three single mutant *Pf*LDH have been solved. The models show that the loss in activity due to oxidation is probably not the result of disulphide bond formation at the subunit interface.

On analysing the three mutant *Pf*LDH, it seems as though the initial enzymatic rate is dependent on the degree to which the dimer subunit has been perturbed.

All three mutant *Pf*LDH structures clearly show electron density for a glycerol molecule at the dimer interface of the tetramer. This is interesting because 10% glycerol is known to stabilise the tetramer.Thus, it may be

possible to further enhance the stabilisation of the oligomeric state of the enzyme by using other small solutes such as sucrose and glucose.

Pig M_4 L-LDH is used as a model for human M_4 L-LDH in antimalarial selectivity studies. In order to overproduce the enzyme, the gene encoding the pig muscle LDH had previously been ligated into a pKK plasmid, this was then transfomed into an *E.coli* strain. On sequencing of the gene 3 sets of non-silent mutations were discovered. This had the effect of mutating 3 residues in the protein. Even though the mutations are at remote sites from the active site, the recombinant triple mutant pig M_4 LDH has a k_{cat} drop of 40% compared to the wild type. The crystal structure of the mutant has been solved. The model shows that the loss in activity of the mutant could probably be attributed to a destabilising change due to the S202N mutation.

References

Adams M.J., Liljas A. and Rossmann M.G. (1973). Functional Anion Binding Sites in Dogfish M_4 Lactate Dehydrogenase. *J. Mol. Biol.*, **76**, 519-531.

Alvarez J.A., Gelpi J.L., Johnsen K., Bernard N., Delcour J., Clarke A.R., Holbrook J.J. and Cortés A. (1997). D-2-Hydroxy-4-methylvalerate Dehydrogenase from *Lactobacillus Delbrueckii* Subsp. *Bulgaricus*. Kinetic Mechanism and pH Dependence of Kinetic Parameters, Coenzyme Binding and Substrate Inhibition. *Eur. J. Biochem.*, **244** , 203-212.

Amemiya Y. and Miyahara J. (1988). Imaging Plate Illuminates Many Fields. *Nature*, **336**, 89-90.

Anderson V.E. and La Reau R.D. (1988). Hydride Transfer Catalysed by Lactate Dehydrogenase Displays Absolute Stereospecificity at C-4 of the Nicotinamide Ring. *J. Am. Chem. Soc.*, **110**, 3695-3697.

Arthur M., Molinas C., Bugg T.D.H., Wright G.D., Walsh C.T. and Courvalin P. (1992). Evidence for *in vivo* Incorporation of D-Lactate into Peptidoglycan Precursors of Vancomycin resistant Enterococci. *Antimicrobial Agents and Chemotherapy*, **36**, 867-869.

Bairoch A. and Apweiler R. (2000). The SWISS-PROT Protein Sequence Database and its supplement TrEMBL in 2000. *Nucl. Acids. Res.*, **28**, 45-48. (Website URL:- http://www.expasy.ch/sprot/sprot-top.html). Wild Type *Lactobacillus Delbrueckii* subsp. *Bulgaricus* D-Hydroxyisocaproate Dehydrogenase Accession I.D.: Q48534. (Masses of selected amino acid residues and atoms in Daltons; His=137.14, Gln=128.14, Met=131.19, Thr=101.11, S=32.064, Se=78.960, Na^+=11, K^+=19).

Baker D.P., Kleanthous C., Keen J.N., Weinhold E. and Fewson C.A. (1992). Mechanistic and Active Site Studies of D-(-)-Mandelic Dehydrogenase from *Rhodotorula graminis*. *Biochem J.*, **281**, 211-218.

Banaszak L.J. and Bradshaw R.A. (1975). Malate Dehydrogenase in *The Enzymes*, **11**, 369-446. (3rd ed., Boyer P.D.) Academic Press.

Barna J.C.J. and Williams D.H. (1984). The Structure and Mode of Action of Glycopeptide Antibiotics of the Vancomycin Group. *Ann. Rev. Microbiol.*, **38**, 339-357.

Benson T.E., Filman D.J., Walsh C.T. and Hogle J.M. (1995). An Enzyme-Substrate Complex involved in Bacterial Cell Wall Biosynthesis. *Nature Struc. Biol.*, **2**, 644-653.

Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N. and Bourne P.E. (2000). The Protein Data Bank. *Nucl. Acids. Res.*, **28**, 235-242. (Website URL:- http://www.rcsb.org/pdb).

Bernard N., Johnsen K., Ferain T., Garmyn D., Hols P., Holbrook J.J. and Delcour J. (1994). NAD^+-Dependent D-2-Hydroxyisocaproate Dehydrogenase of *Lactobacillus Delbrueckii* Subsp. *Bulgaricus*. Gene Cloning and Characterization. *Eur. J. Biochem.*, **224** , 439-446.

Bernard N., Johnsen K., Gelpi J.L., Alvarez J.A., Ferain T., Garmyn D., Hols P., Cortés A., Clarke A.R., Holbrook J.J. and Delcour J. (1996). D-2-Hydroxyisocaproate Dehydrogenase from *Lactobacillus Delbrueckii* Subsp. *Bulgaricus*. Mutagenic Analysis of Catalytically Important Residues and a Mutant with Enhanced Potential for Chemoenzymic Conversions. *J. Biol. Chem.*, **270** , ?.

Bernard N., Johnsen K., Holbrook J.J. and Delcour J. (1995). D175 Discriminates between NADH and NADPH in the Coenzyme Binding Site of *Lactobacillus*

Delbrueckii Subsp. *Bulgaricus* D-Lactate Dehydrogenase. *Biochem. Biophys. Res. Commun.*, **208 (3)**, 895-900.

Birktoft J.J., Fernley R.T., Bradshaw R.A. and Banaszak L.J. (1982). Molecular Structure and Biological Activity. (1st ed.) Elsevier.

Blow D.M. (1995). Introduction to Rotation and Translation Functions. *Molecular Replacement Proceedings of the Daresbury Study Weekend*, 2-7.

Blundell T.L. and Johnson L.N. (1976). Protein Crystallography. London: Academic Press.

Boernke W.E., Millard C.S., Stevens P.W., Kakar S.N., Stevens S.J. and Donnely M.I. (1995). Stringency of Substrate Specificity of *Escherichia coli* Malate Dehydrogenase. *Arch. Biochem. Biophys.*, **322**, 43-52.

Bradford M.M. (1976). A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Anal. Biochem.*, **72**, 248-254.

Brock T.D. and Madigan M.T. (1988). Biology of Microorganisms. (5th ed.) New Jersey: Prentice Hall International, Inc.

Brünger A.T. Kuriyan J. and Karplus M.(1987). Crystallographic R-factor Refinement by Molecular Dynamics. *Science*, **235**, 458-460.

Brünger A.T. (1992a). XPLOR Version 3.1. Yale University Press.

Brünger A.T. (1992b). The Free R- Value: A Novel Statistical Quantity for Assessing the Accuracy of Crystal Structures. *Nature*, **355**, 472-474.

Brünger A.T., Adams P.D., Clore G.M., DeLano W.L., Gros P., Grosse-Kunstleve R.W., Jiang J.-S., Kuszewski J., Nilges N., Pannu N.S., Read R.J., Rice L.M.,

Simonson T. and Warren G.L. (1998). Crystallography and NMR System (CNS): A New Software System for Macromolecular Structure Determination. *Acta Cryst. D*, **54**, 905-921.

Carroll B. and Behrens R. (1998). Malaria: A Changing Pattern of Resistance. *Practitioner*, **242**, 349-365.

Casazza J.P., Felver M.E. and Veech R.L. (1984).The Metabolism of Acetone in Rat. *J. Biol. Chem.*, **259**, 231-236.

Cedergren-Zepperzauer E., Samama J-P. and Eklund H. (1982). Crystal Structure Determinations of Coenzyme Analogue and Substrate Complexes of Liver Alcohol Dehydrogenase: Binding of 1,4,5,6-Tetrahydronicotinamide Adenine Dinucleotide and *trans*-4-(N,N-Dimethylamino)cinnamaldehyde to the enzyme. *Biochemistry*, **21**, 4895-4908.

Chait B.T. (1994). Mass Spectrometry - A Useful for the Protein Crystallographer and NMR Spectroscopist. *Structure*, **2**, 465-467.

Chait B.T. and Kent S.B.H. (1992). Weighing Naked Proteins: Practical, High-Accuracy Mass Measurement of Peptides and Proteins. *Science*, **257**, 1885-1894.

Chapman A.D.M., Cortés A., Daffron T.R., Clarke A.R. and Brady R.L. (1999). Structural Basis of Substrate Specificity in Malate Dehydrogenases: Crystal Structure of a Ternary Complex of Porcine Cytoplasmic Malate Dehydrogenase, α-Ketomalonate and TetrahydroNAD. *J. Mol. Biol.*, **285**, 703-712.

Chothia C. and Lesk A.M. (1986). The Relation Between the Divergence of Sequence and Structure in Proteins. *EMBO J.*, 5, 823-826.

Clarke A.R. and Daffron T.R. (1998). NAD^+ and NADP-Linked Reactions in *Comprehensive Biological Catalysts*. Academic Press, 1-80.

Clarke A.R., Wigley D.B., Chia W.N., Barstow D., Atkinson T. and Holbrook J.J. (1986). Site-Directed Mutagenesis Reveals Role of Mobile Arginine Residue in Lactate Dehydrogenase. *Nature*, **324**, 699-702.

Collaborative Computational Project, Number 4 (1994). The CCP4 Suite - Programs for Protein Crystallography. *Acta Cryst. D*, **50**, 760-763.

Corpet F. (1988). Multiple Sequence Alignment with Hierarchical Clustering. *Nucl. Acids. Res.*, **22**, 10881-10890.

Cowie D.B. and Cohen G.N. (1957). Biosynthesis by *Escherichia Coli* of Active Altered Proteins Containing Selenium instead of Sulphur. *Biochimica et Biophysica Acta*, **26**, 252-261.

Dalziel K. (1975). Kinetics and Mechanism of Nicotinamide-Nucleotide-Linked Dehydrogenases in *The Enzymes*, **11**, 1-60. (3rd ed., Boyer P.D.) Academic Press.

Dave K.G., Dunlap R.B., Jain M.K. Cordes E.H. and Wenkert E. (1968).Hydrogenation of NAD^+. *J. Biol. Chem.*, **243**, 1073-1074.

Dengler U., Niefind K., Kieß M. and Schomburg D. (1997). Crystal Structure of a Ternary Complex of D-2-Hydroxyisocaproate Dehydrogenase from *Lactobacillus casei*, NAD^+ and 2-Oxoisocaproate at 1.9 Å Resolution. *J. Mol. Biol.*, **267**, 640-660.

Dixon M. (1955). Lactic Dehydrogenase from Yeast. *Methods Enzymol.*, **1**, 444-449.

Dodson E.J. (1995). Introduction to Rotation and Translation Functions. *Molecular Replacement Proceedings of the Daresbury Study Weekend.*

Dodson E.J., Gover S. and Wolf W. (eds) (1992). Molecular Replacement. *Molecular Replacement Proceedings of the CCP4 Daresbury Study Weekend.*

Doublié S. (1997). Preparation of Selenomethionyl Proteins for Phase Determination. *Methods Enzymol.*, **276**, [29] 523-530.

Doublié S., Kapp U., Åberg A., Brown K., Strub K. and Cusack S. (1996). Crystallization and Preliminary X-ray analysis of the 9 kDa Protein of the Mouse Signal Recognition Particle and the Selenomethionyl-SRP9. *FEBS Lett.*, **384**, 219-221.

Douglas K.T. and Shinkai S. (1985). Chemical Basis of the Action of Glyoxalase I, an Anticancer Target Enzyme. *Angew. Chem. Int. Ed. Engl.*, **24**, 31-44.

Dreissen H.P.C. and Tickle I.J. (1996). *Methods in Molecular Biology: Crystallography Methods and Protocols*. (1st ed., Jones C., Mulley B. and Sanderson M.R.) New York: Humana Press.

Drenth J. (1995). Principles of Protein X-ray Crystallography. (2nd ed.) Springer-Verlag.

Dunn C.R. (1989). A Crystallographic Study of Structural Changes in L-Lactate Dehydrogenase Induced by the Binding of Substrate. Ph.D. Thesis, Bristol University.

Dunn C.R., Banfield M.J., Barker J.J., Higham C.W., Moreton K.M., Turgut-Balik D., Brady R.L. and Holbrook J.J. (1996). The Structure of Lactate Dehydrogenase from *Plasmodium Falciparum* Reveals a New Target for Anti-Malarial Design. *Nature Struc. Biol.*, **3**, 912-915.

Dunn C.R. and Holbrook J.J.(1996). Crystal Structure of *Lactobacillus Helveticus* D-Lactate Dehydrogenase Complexed with NADH and Oxamate. PDB I.D.: 2DLD.

Dunn C.R., Holbrook J.J.and Muirhead H.(1991). Crystal Structure of *Sus Scrofa* (Porcine/Pig) M_4 L-Lactate Dehydrogenase Complexed with NADH and Oxamate. PDB I.D.: 9LDT.

Dym O., Pratt E.A., Ho. C. and Eisenberg D. (2000). The Crystal Structure of D-Lactate Dehydrogenase, a Peripheral Membrane Respiratory Enzyme. *Proc. Nat. Acad. Sci. USA*, **97**, 9413-9429.

El Harwani A.S., Sessions R.B., Moreton K.M. and Holbrook J.J. (1996). Guided Evolution of Enzymes with New Substrate Specificities. *J. Mol. Biol.*, **264**, 97-110.

Engh R.A. and Huber R. (1991). Accurate Bond and Angle Parameters for X-ray Protein Structure Refinement. *Acta Cryst. A*, **47**, 392-400..

Evans G. and Pettifer R.F. (2001). CHOOCH: A Program for Deriving Anomalous-Scattering Factors from X-ray Fluorescence Spectra. *J. Appl. Cryst.*, **34**, 82-86.

ExPASy-Tools program (2001). Compute pI / Mw for SWISS-PROT / TrEMBL Entries or a User-Entered Sequence. (Website URL:- http://www.expasy.cbr.nrc.ca/tools/pi_tools.html).

Fenn B.J., Mann M., Meng C.K., Wong S.H. and Whitehouse C.M. (1989). Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science*, **246**, 64-71.

Feil I., Lerch H.-P. and Schomburg D. (1994). Deletion Variants of L-2-Hydroxyisocaproate Dehydrogenase. Probing Substrate Specificity. *Eur. J. Biochem.*, **223**, 857-863.

Garman E.F. and Schneider T.R. (1997). Macromolecular Cryocrystallography. *J. Appl. Cryst.*, **30**, 211-237.

Gerstein M. and Chothia C. (1991). Analysis of Protein Loop Closure. Two Types of Hinges Produce One Motion in Lactate Dehydrogenase. *J. Mol. Biol.*, **220**, 133-149.

Goldberg J.D., Yoshida T. and Brick P. (1994). Crystal Structure of a NAD-dependent D-Glycerate Dehydrogenase at 2.4Å Resolution. *J. Mol. Biol.*, **236**, 1123-1140.

Gomez M.S., Piper R.C., Hunsaker L.A. Royer R.E., Deck L.M., Makler M.T. and Vander Jagt D.L. (1997). Substrate and Cofactor Specificity and Selective Inhibition of Lactate Dehydrogenase from the Malarial Parasite *P.falciparum*. *Mol. Biochem. Parasit.*, **90**, 235-246.

Grant G.A. (1989). A New Family of 2-Hydroxyacid Dehydrogenases. *Biochem. Biophys. Res. Commun.*, **165 (3)**, 1371-1374.

Grau U.M., Trommer W.E. and Rossmann M.G. (1981). Structure of the Active Ternary Complex of Pig Heart Lactate Dehydrogenase with *S*-lac-NAD at 2.7Å Resolution. *J. Mol. Biol.*, **151**, 289-307.

Greenler J.M., Sloan J.S., Schwartz B.W. and Becker W.M. (1989). Isolation, Characterization and Sequence Analysis of a Full-Length cDNA Clone Encoding NADH-Dependent Hydroxypyruvate Reductase from Cucumber. *Plant Mol. Biol.*, **13**, 139-150.

Hackert M.L., Ford G.C. and Rossmann M.G. (1973).Molecular Orientation and Position of the Pig M_4 and H_4 Isoenzymes of Lactate Dehydrogenase in their Crystal Cells. *J. Mol. Biol.*, **78**, 665-673.

Hahn T. (editor) (1995). International Tables for Crystallography Volume A. (4th Edition) Kluwer Academic Publishers.

Hall D.R., Jacobsen M.P. and Winzor D.J. (1995). Stabilizing Effect of Sucrose Against Irreversible Denaturation of Rabbit Muscle Lactate Dehydrogenase. *Biophysical Chemistry*, **57**, 47-54.

Harker D. (1956). The Determination of the Phases of the Structure Factors of Non-Centrosymmetric Crystals by the Method of Double Isomorphous Replacement. *Acta Cryst.*, **9**, 1-9.

Hart K.W., Clarke A.R., Wigley D.B., Waldman A.D.B., Chia W.N., Barstow D.A., Atkinson T., Jones J.B. and Holbrook J.J. (1987). A Strong Carboxylate-Arginine Interaction is Important in Substrate Orientation and Recognition in Lactate-Dehydrogenase. *Biochimica et Biophysica Acta*, **914**, 294-298.

Helliwell J.R.. (1992). Macromolecular Crystallography with Synchrotron Radiation. (1st ed.) Cambridge University Press.

Henderson R. (1990). Cryo-Protection of Protein Crystals Against Radiation Damage in Electron and X-ray Diffraction. *Proc. R. Soc. London Ser. B.*, **241**, 6-8.

Hendrickson W.A. (1991). Determination of Macromolecular Structures from Anomalous Diffraction of Synchrotron Radiation. *Science*, **254**, 51-58.

Hendrickson W.A., Horton J.R. and LeMaster D.M. (1990). Selenomethionyl Proteins Produced for Analysis by Multiwavelength Anomalous Diffraction (MAD): A Vehicle for Direct Determination of Three-Dimensional Structure. *EMBO J.*, 9, 1665-1672.

Hendrickson W.A., Smith J.L., Phizackerley R.P. and Merritt E.A. (1988).Crystallographic Structure Analysis of Lamprey Hemoglobin From Anomalous Dispersion of Synchrotron Radiation. *Proteins*, **4**, 77-88.

Holbrook J.J., Liljas A., Steindel S.J. and Rossmann M.G. (1975). Lactate Dehydrogenase in *The Enzymes*, **11**, 191-292. (3rd ed., Boyer P.D.) Academic Press.

Huber R.E., Segel I.H and Criddle R.S. (1967). Growth of *Escherichia Coli* on Selenate. *Biochimica et Biophysica Acta*, **141**, 573-586.

Hummel W. and Kula M-R. (1989). Dehydrogenases for the Synthesis of Chiral Compounds. *Eur. J. Biochem.*, **184**, 1-13.

Hummel W., Schütte H. and Kula M-R. (1988). D-(-)-Mandelic Dehydrogenase of *Lactobacillus curvatus*. *Appl. Microbiol. Biotechnol.*, **28**, 433-439.

Hummel W., Schütte H. and Kula M-R. (1985). D-2-Hydroxyisocaproate Dehydrogenase of *Lactobacillus casei*. A New Enzyme Suitable for Stereospecific Reduction of 2-Ketocarbxylic Acids. *Appl. Microbiol. Biotechnol.*, **21**, 7-15.

Jacarik J. and Kim S.H. (1991). Sparse-Matrix Sampling - A screening Method for Crystallisation of Proteins. *J. Appl.Cryst.*, **24**, 409-411.

James R.W. (1948). The Optical Principles of the Diffraction of X-rays. London: Bell Press.

Jecsai G. (1962). *Acta Physiol. Hung.*, **20**, 339.

Jensen R.A. (1976). Enzyme Recruitment in Evolution of New Function. *Ann. Rev. Microbiol.*, **30**, 409-425.

Johnsen K. (1995). D-2-Hydroxyacid Dehydrogenases. Ph.D. Thesis, Bristol University.

Jones T.A., Zou J.-Y., Cowan S.W. and Kjeldgaard M. (1991). Improved Methods for Building Protein Models in Electron Density Maps and the Location of Errors in these Models. *Acta Cryst. A*, **47**, 110-119.

Keller P.A., Henrick K., McNeil P., Moodie S. and Barton G.J. (1998). Deposition of Macromolecular Structures. *Acta Cryst. D*, **54**, 1105-1108. (Website URL:- http://msd.ebi.ac.uk).

Kendrew J.C., Bodo G. Dintzis, H.M., Parrish R.G., Wyckoff H. and Phillips D.C. (1958). A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-ray Analysis. *Nature (London)*, **181**, 662-666.

Kissinger C.R., Gehlhaar D.K. and Fogel D.B. (1999). Rapid Automated Molecular Replacement by Evolutionary Search. *Acta Cryst. D*, **55**, 484-491.

Kleywegt G.J. (2000). Uppsala Software Factory-Software and Resources for Macromolecular Crystallography and Structural Biology. (Website URL:- http://alpha2.bmc.uu.se/~gerard/manuals/gerard_manuals.html).

Knell A.J. (1991). Malaria. Oxford University Press, Oxford.

Kochhar S., Lamzin V.S., Razeto A. Delley M., Hottinger H. and Germond J.-E.(2000). Roles of His205, His296, His303 and Asp259 in Catalysis by NAD^+-Specific D-Lactate Dehydrogenase. *Eur. J. Biochem.*, **267** , 1633-1639.

Kochhar S., Hunziker P.E, Leong-Morgenthaler P. and Hottinger H. (1992). Primary Structure, Physiochemical Properties, and Chemical Modification of NAD^+-dependent D-Lactate Dehydrogenase. *J. Biol. Chem.*, **267**, 8499-8513.

Kornberg A. (1955). Lactic Dehydrogenase of Muscle. *Methods In Enzymology*, **1**, 441-443.

Kraulis P.J. (1991). MOLSCRIPT: A Program to Produce both Detailed and Schematic Plots of Protein Structures. *J. Appl. Cryst.*, **24**, 946-950.

Kraut J. (1977). Serine Proteases: Structure and Mechanism of Catalysis. *Annu. Rev. Biochem.*, **46**, 331-358.

Kronig R. and Kramers H.A. (1928). *Z. Phys.*, **48**, 174.

Laskowski R.A., MacArthur M.W., Moss D.S. and Thornton J.M. (1993). PROCHECK - A Program to Check the Stereochemical Quality of Protein Structures. *J. Appl. Cryst.*, **26**, 283-291.

Long G.L. and Kaplan N.O. (1968). D-Lactate Specific Pyridine Nucleotide Lactate Dehydrogenase in Animals. *Science*, **162**, 685-686.

Makler M.T. and Hinrichs D.J. (1993). Measurement of the Lactate Dehydrogenase activity from *Plasmodium falciparum* as an Assessment of Parasitemia. *American Journal of Tropical Medicine and Hygiene*, **48**, 205-210.

Makman R.S. and Sutherland E.W. (1965).Adenosine 3′,5′-Phosphate in *Escherichia Coli. J. Biol. Chem.*, **240**,1309-1314.

Matthews B.W. (1968). Solvent Content of Protein Crystals. *J. Mol. Biol.*, **33**, 491-497.

McPherson A. (1990). Current Approaches to Macromolecular Crystallization. *Eur. J. Biochem.*, **189**, 1-23.

McPherson A. (1973). Binding of Oxamate to the Apoenzyme of Dogfish M_4 Lactate Dehydrogenase. *J. Mol. Biol.*, **76**, 528-531.

McRee D.E. (1999). XtalView/Xfit - A Versatile Program for Manipulating Atomic Coordinates and Electron Density. *J. Structural Biology*, **125**, 156-165.

Murshudov G.N., Vagin A.A., Lebedev A., Wilson K.S. and Dodson E.J. (1999). Efficient Anisotropic Refinement of Macromolecular Structures using FFT. *Acta Cryst. D*, **55**, 247-255.

Navaza J. (1994). AMoRe - An automated Package for Molecular Replacement. *Acta Cryst. A*, **50**, 157-163.

Niefind K., Hecht H.-J. and Schomburg D. (1995). Crystal Structure of L-2-Hydroxyisocaproate Dehydrogenase from *Lactobacillus confusus* at 2.2Å Resolution. An Example of Strong Asymmetry Between Subunits. *J. Mol. Biol.*, **251**, 256-281.

Noyes B.E. and Bradshaw R.A. (1973).L-3-Hydroxyacyl Coenzyme A Dehydrogenase from Pig Heart Muscle. *J. Biol. Chem.*, **248**, 3052-3059.

Olliaro and Yuthavong (1999). An Overview of Chemotherapeutic Targets for Antimalarial Drug Discovery. *Pharmacol. Ther.*, **81**, 99-110.

Otwinowski, Z. and Minor W. (1996). Processing of X-ray Diffraction Data collected in Ocillation Mode. *Methods Enzymol.*, **276**, [29] 307-326.

Parker D.M. and Holbrook J.J. (1977). An Oil-Water Histidine Mechanism for the Activation of Coenzyme in the α-Hydroxyacid Dehydrogenases. Pyridine Nucleotide-Dependent Dehydogenases (1st ed., Sund H.) Walter de Grunyter & Co., Berlin.

Pflugrath J.W. (1992). Developments in X-ray Detectors. *Current Opinion in Structural Biology*, **2**, 811-815.

Ramachandran N.G. and Sasisekharen V. (1968). Conformation of Polypeptides and Proteins. *Advances in Protein Chemistry*, **23**, 283-437.

Rao S.T. and Rossmann M.G. (1973). Comparison of Super-Secondary Structures in Proteins. *J. Mol. Biol.*, **76**, 241-256.

Read J.A., Wilkinson K.W., Tranter R., Sessions R.B. and Brady R.L. (1999).Chloroquine Binds in the Cofactor Binding Site of *Plasmodium falciparum* Lactate Dehydrogenase. *J. Biol. Chem.*, **274**, 10213-10218.

Read J.A., Winter V.J., Eszes C.M., Sessions R.B. and Brady R.L. Structural Basis for Altered Activity of M- and H-Isozyme Forms of Human Lactate Dehydrogenase. *(Submitted for Publication).*

Rhodes G. (1993). Crystallography Made Crystal Clear. (2nd ed) Academic Press.

Rosenblum I.Y., Antkowiak D.H., Sallach H.J., Flanders L.E. and Fahien L.A. (1971). Purification and Regulatory Properties of Beef Liver D-Glycerate Dehydrogenase. *Arch. Biochem. Biophys.*, **144**, 375-383.

Rossmann M.G. and Blow D.M. (1962). The Detection of Sub-Units Within the Crystallographic Sub-Unit. *Acta Cryst.*, **15**, 18-23.

Rossmann M.G., Adams M.J., Buehner M., Ford G.C., Hackert M.L., Lentz P.J., McPherson A., Schevitz R.W. and Smiley J.E. (1971). *Cold Spring Harbour Symposia on Quantitative Biology*, **36**, 179-191.

Rossmann M.G., Moras D. and Olsen K.W. (1974). Chemical and Biological Evolution of a Nucleotide-Binding Protein. *Nature*, **250**, 194-199.

Rossmann M.G., Liljas A., Brändén C-I. and Banaszak L.J. (1975). Evolutionary and Structural Relationships among Dehydrogenases in *The Enzymes*, **11**, 61-102. (3rd ed., Boyer P.D.) Academic Press.

Roth M. (1968). Fluorimetric Assay of Enzymes. *Methods of Biochemical Analysis*, **17**, 189-277.

Royer R.E., Deck L.M., Campos N.M., Hunsaker L.A. and Vander Jagt D.L. (1986). Biologically Active Derivatives of Gossypol- Synthesis and Anti-Malarial Activies of Peri-Acylated Gossylic Nitriles. *J.Med.Chem.*, **29**, 1799-1801.

Rupp B. (2000). 'Crystallography 101' (Website URL:- http://www-structure.llnl.gov/Xray).

Sakowicz R., Gold M. and Jones J.B. (1995). Partial Reversal of the Substrate Stereospecificity of an L-Lactate Dehydrogenase by Site-Directed Mutagenesis. *J. Am. Chem. Soc.*, **117**, 2387-2394.

Sambrook J., Fritsche E.T. and Maniatis T. (1989). Molecular Cloning, a Laboratory Manual. New York: Cold Spring Harbor Press.

Schoenlein P.V. Roa B.B. and Winkler M.E. (1989). Divergent Transcription of *pdxB* and Homology Between the *pdxB* and *serA* Gene-Products in *Escherichia coli* K-12. *J. Bacteriol.*, **177**, 6084-6092.

Schuller D.J., Grant G.A. and Banaszak L.J. (1995). The Allosteric Ligand Site in the V_{max}-Type Cooperative Enzyme Phosphoglycerate Dehydrogenase. *Nature Struc. Biol.*, **2**, 69-76.

Schütte H., Hummel W. and Kula M.-R. (1984). L-2-Hydroxyisocaproate Dehydrogenase- A New Enzyme from *Lactobacillus confusus* for the stereospecific reduction of 2-ketocarboxylic acids. *Appl. Microbiol. Biotechnol.*, **19**, 167-176.

Schwartz M., O'Donnell J. and Sofer W. (1979). Origin of the Multiple Forms of Alcohol Dehydrogenase from *Drosophila melanogaster*. *Arch. Biochem. Biophys.*, **194**, 365-378.

Sessions R.B., Dewar V., Clarke A.R. and Holbrook J.J. (1997). A model of *Plasmodium falciparum* Lactate Dehydrogenase and its Implications for the Design of Improved Antimalarials and the Enhanced Detection of Parasitaemia. *Protein Engineering*, **10**, 301-306.

Shoemark D.K. (2000). Kinetic Characterization of the Lactate Dehydrogenase from *Plasmodium Falciparum*. Ph.D. Thesis, Bristol University.

Siebanaller J.F., Orr T.L., Olwin B.B., and Taylor S.S. (1983). Comparison of the D-Lactate Stereospecific Dehydrogenase of *Limulus Polyphemus* with Active Site Regions of L-Lactate Dehydrogenases. *Biochimica et Biophysica Acta*, **749**, 153-162.

Smith J.L. (1991). Determination of Three-Dimensional Structure by Multiwavelength Anomalous Diffraction. *Current Opinion in Structural Biology*, **1**, 1002-1011.

Stoll V.S and Blanchard J.S. (1990). Buffers: Principles and Practice. *Methods Enzymol.*, **182**, 24-38.

Thevenot D.R., Godinot C. and Gautheron D.C. (1975). Binding of L-Glutamate to Glutamate Dehydrogenase in the Presence of 1,4,5,6-Tetrahydronicotinamide Adenine Dinucleotide. *FEBS Letters*, **54**, 206-211.

Tickle I.J. (1998). Molecular Replacement, Bath Protein Crystallography Summer School Lectures.

Tobey K.L. and Grant G.A. (1986). The Nucleotide-Sequence of the *serA* gene of *Escherichia coli* and the Amino-Acid Sequence of the Encoded Protein, D-3-Phosphoglycerate Dehydrogenase. *J. Biol. Chem.*, **261**, 2179-2183.

Turner P.C, McLennan A.G., Bates A.D. and White M.R.H. (1997). Instant Notes in Molecular Biology. (1st ed., Robinson R.) BIOS Scientific Publishers.

Vander Jagt D.L., Hunsaker L.A. and Heidrich J.E. (1981). Partial Purification and Characterisation of Lactate Dehydrogenase from *Plasmodium falciparum. Mol. Biochem. Parasit.*, **4**, 255-264.

Van Duyne G.D., Standaert R.F., Karplus P.A., Schreiber S.L. and Clardy J. (1993). Atomic Structures of the Human Immunophilin FKBP-12 Complexes with FK506 and Rapamycin. *J. Mol. Biol.*, **229**, 105-124.

Vinals C., De Bolle X., Depiereux E. and Feytmans E. (1995). Knowledge-based Modelling of the D-Lactate Dehydrogenase Three Dimensional Structure. *Proteins*, **21**, 307-316.

Vinals C., Depiereux E. and Feytmans E. (1995). Predication of Structurally Conserved Regions of D-Specific Hydroxy Acid Dehydrogenase by Multiple Alignment with Formate Dehydrogenase. *Biochem. Biophys. Res. Commun.*, **192 (1)**, 182-188.

Voet D. and Voet J.G. (1995). Biochemistry. (2nd ed.) John Wiley & Sons, Inc.

Waldman A.D.B., Birdsall B., Roberts G.C.K. and Holbrook J.J. (1987). The Use of Site-Directed Mutagenesis and Time-Resolved Fluorescence Spectroscopy to Assign the Fluorescence Contributions of Individual Tryptophan Residues in *Bacillus-stearothermophilus* Lactate-Dehydrogenae. *Biochimica et Biophysica Acta*, **914**, 294-298.

Watson J.D., Gilman M., Witkowski J. and Zoller M. (1992). Recombinant DNA. (2nd ed.) Scientific American Books.

Weinhold E.G., Glasfled A., Ellington A.D. and Benner S.A. (1991). Structural Determinants of Stereospecificity in Yeast Alcohol Dehydrogenase. *Proc. Natl. Acad. Sci.*, **88**, 8420-8424.

White J.L., Hackert M.L., Buehner M., Adams M.J., Ford G.C., Lentz P.J., Smiley J.E., Steindel S.J. and Rossmann M.G. (1976). A Comparison of the Structures of Apo Dogfish M_4 Lactate Dehydrogenase and its Ternary Complex. *J. Mol. Biol.*, **102**, 759-779.

WHO (1987). The Biology of Malaria Parasites. World Health Orginisation. Geneva

WHO (1996). The World Health Report 1996, Fighting Disease Fostering Development. World Health Orginisation. Geneva

Wigley D.B., Gamblin S.J., Turkenburg J.P., Dodson E.J., Piontek K., Muirhead H.M. and Holbrook J.J. (1992). Structure of a Ternary Complex of an Allosteric Lactate Dehydrogenase from Bacillus stearothermophilus at 2.5Å Resolution. *J. Mol. Biol.*, **223**, 317-335.

Wilks H.M., Halsall D.J., Atkinson T., Chia W.N, Clarke A.R. and Holbrook J.J. (1990). Designs for a Broad Substrate Specificity Keto Acid Dehydrogenase. *Biochemistry*, **29**, 8587-8591.

Wilks H.M., Moreton K.M., Halsall D.J., Hart K.W., Sessions R.B., Clarke A.R. and Holbrook J.J. (1992). Design of a Specific Phenyllactate Dehydrogenase by Peptide Loop Exchange on the *Bacillus stearothermophilus* Lactate Dehydrogenase Framework. *Biochemistry*, **31**, 7802-7806.

Wilks H.M., Hart K.W., Feeney R., Dunn C.R., Muirhead H., Chia W.N., Barstow D.A., Atkinson T., Clarke A.R. and Holbrook J.J. (1988). A Specific, Highly Active Malate Dehydrogenase by Redesign of a Lactate Dehydrogenase Framework. *Science*, **242**, 1541-1544.

Wilson A.J.C. (1942). Determination of Absolute from Relative X-Ray Intensity Data. *Nature*, **150**, 151-152.

Winston R.L. and Fitzgerald M.C. (1998). Concentration and Desalting of Protein Samples for Mass Spectrometry Analysis. *Analytical Biochemistry*, **262**, 83-85.

Wolf W., Evans P.R. and Leslie A.G.W. (eds.) (1991). Isomorphous Replacement and Anomalous Scattering. CCP4 Daresbury Study Weekend: Daresbury Laboratory, Warrington.

Woolfson M.M. (1997). An Introduction to X-ray Crystallography. (2nd ed.) Cambridge University Press.

Wu Z., Wright G.D. and Walsh C.T. (1995). Overexpression, Purification, and Characterization of *vanX*, a D-, D- dipeptidase which is Essential for Vancomycin Resistance in *Enterococcus faecium* BM4147. *Biochemistry*, **34**, 2455-2463.

Yang W., Hendrickson W.A., Kalman E.T. and Crouch R.J. (1990).Expression, Purification, and Crystallization of Natural and Selenomethionyl Recombinant Ribonuclease H from *Escherichia coli*. *J. Biol. Chem.*, **265**, 13553-13559.